



Benemérita Universidad Autónoma de Puebla

Facultad de Ciencias Físico Matemáticas

Semántica en mamografías con deep Learning: Eficiencia de
CNN modificadas con DeepLabV3+

Tesis presentada al

Colegio de Física

como requisito parcial para la obtención del grado de

LICENCIADO EN FÍSICA APLICADA

por

Luis Gerardo Hernández Román

Asesorado por

Dr. Cristian Heber Zepeda Fernández

Puebla Pue.
17 de octubre de 2025



Benemérita Universidad Autónoma de Puebla

Facultad de Ciencias Físico Matemáticas

Semántica en mamografías con deep Learning: Eficiencia de
CNN modificadas con DeepLabV3+

Tesis presentada al

Colegio de Física

como requisito parcial para la obtención del grado de

LICENCIADO EN FÍSICA APLICADA

por

Luis Gerardo Hernández Román

Asesorado por

Dr. Cristian Heber Zepeda Fernández

Puebla Pue.
17 de octubre de 2025

Título: Semántica en mamografías con deep Learning: Eficiencia de CNN modificadas con DeepLabV3+

Estudiante: LUIS GERARDO HERNÁNDEZ ROMÁN

COMITÉ

Dr. Lucio Fidel Rebolledo Herrera
Presidente

Dr. Juan Moisés Arredondo Velázquez
Secretario

Dr. Eduardo Moreno Barbosa
Vocal

Dr. Jorge Velázquez Castro
Vocal

Dr. Cristian Heber Zepeda Fernández
Asesor

Agradecimientos y dedicatorias

A mis padres, José Luis y Carmela Román, quienes con su amor incondicional, sacrificios y sabios consejos han sido la base de este logro. Gracias por enseñarme con el ejemplo el valor del esfuerzo, la perseverancia y la humildad. Este logro no sería posible sin su apoyo constante, su paciencia y el cariño con el que me acompañaron en cada paso de este camino. Todo lo que hoy celebro es también un reflejo de lo que me han dado y del inmenso orgullo que siento de ser su hijo.

A mis sobrinos Alan Daniel, Evelyn y Mateo, quienes con su alegría iluminan mi vida y me recuerdan la importancia de sonreír incluso en los momentos difíciles. Ustedes han sido una fuente constante de motivación y felicidad en este camino.

A mi hermano Daniel Hernández y su esposa Esmeralda, y a mi hermano Eduardo Hernández, por su apoyo incondicional y por estar siempre a mi lado, brindándome ánimo, confianza y cariño en cada paso de este camino. Su respaldo ha sido fundamental para alcanzar este logro.

A mi querida amiga Ariadna Hernández, mi comadre y confidente, a quien le tengo un profundo cariño y admiración. Gracias por estar siempre a mi lado, por escucharme, animarme en los momentos difíciles y celebrar conmigo cada logro. Tu amistad ha sido un regalo invaluable en mi vida, y este logro también te pertenece.

A mi amigo Andrés Grijalva, mi mejor compañero en este camino universitario. Gracias por tu amistad sincera, por las risas compartidas en los momentos de cansancio y por estar siempre dispuesto a ofrecer tu apoyo. Tu compañía hizo de este recorrido una experiencia más ligera y significativa, y siempre guardaré con cariño los recuerdos que construimos juntos.

A mis amigos Iran Melani, Javier Marcial, Karla Edith, Josué Cruz, Enrique Martínez y Fernanda Pamela, por su amistad y compañía. Gracias por cada salida, por el tiempo compartido y por todos los buenos momentos que hicieron de esta etapa una experiencia inolvidable. Su presencia marcó mi vida universitaria y siempre llevaré con cariño los recuerdos que construimos juntos.

A mis amigos Jonathan, Sandra Susana y Nancy, quienes siempre han estado presentes de corazón. Su amistad sincera y su apoyo constante han sido un pilar importante en mi vida, recordándome que los lazos verdaderos perduran más allá del tiempo y la distancia.

Expreso mi más sincero agradecimiento al Dr. Cristian Heber, por su confianza, orientación y valioso apoyo a lo largo de todo este trabajo. Su guía, paciencia y motivación fueron fundamentales para la culminación de esta tesis, y me siento profundamente agradecido por haber contado con su experiencia y acompañamiento en esta etapa tan importante de mi formación académica.

Agradezco también al Dr. Juan Castillo y al Dr. Juan Moisés, por su orientación y ayuda en el aprendizaje del entorno MATLAB, conocimientos que fueron fundamentales en el desarrollo de esta investigación.

Al Dr. Lucio Fidel, Dr. Juan Moisés, Dr. Eduardo y al Dr. Jorge por acceder a formar parte del mi jurado de tesis. Por su orientación en los pasos finales para mejorar este trabajo.

Finalmente, a todos mis profesores, quienes a lo largo de estos cuatro años compartieron sus enseñanzas y contribuyeron de manera invaluable a mi formación académica y personal.

Índice general

1. Introducción	1
2. Física y Medicina	5
2.1. Anatomía	5
2.1.1. Sistema de flujo linfático	5
2.1.2. Estructura de la mama y las glándulas mamarias	6
2.1.3. Patrones mamográficos	6
2.2. Cáncer en la Mama	7
2.2.1. Clasificación radiológica	8
2.3. Sistema de imagen mamográfico	9
2.3.1. Generador de alta frecuencia	10
2.3.2. Tubo de rayos X	11
2.3.3. Diana del tubo de rayos x	13
2.3.4. Punto focal	15
2.3.5. Interacción de los Rayos X con la Mama	15
2.3.6. Captura de la imagen	17
3. Estado del arte	19
3.1. Introducción	19
3.2. Neuronas	19
3.3. El Modelo multicapa	20
3.4. Redes neuronales convolucionales	22
3.4.1. Resnet y los bloque residuales	23
3.4.2. Resnet- 50	24
3.5. Segmentación semántica	25
3.5.1. DeepLabV3+	25
3.6. Función de Pérdida	27
3.6.1. Regularización L2	27
3.6.2. Descenso del gradiente	28
3.6.3. Tasa de Aprendizaje inicial (ILR)	28
4. Metodología	29
4.1. Software	29
4.2. Diagrama de flujo	29
4.3. Base de datos	31
4.4. Preprocesamiento de las imágenes	31
4.4.1. Configuración de las nuevas capas	34
4.5. Entrenamiento de las CNNs	35

5. Resultados	37
5.1. Eficiencia de las CNNs	37
5.1.1. Configuración de las pruebas	37
5.2. Curvas de Pérdida del entrenamiento	37
5.3. Métricas	40
5.3.1. Precisión (Accuracy)	40
5.3.2. Intersección sobre la unión (IoU)	41
5.3.3. Resultados globales	44
6. Discusión y conclusiones	47
A.	49
A.1. Preprocesamiento y entrenamiento	49
A.2. Funciones	51
Bibliografía	53

Capítulo 1

Introducción

El cáncer de mama es una de las principales causas de muerte en México y siendo la principal causa de muerte por cáncer en mujeres mayores de 25 años. Se trata de una enfermedad que afecta mayormente a mujeres aunque también puede presentarse en hombres. La detección temprana es fundamental para mejorar la tasa de supervivencia, pero existen retos significativos en la conciencia y el acceso a servicios de salud eficientes. Cuando el cáncer de mama se identifica en la etapa I, la tasa de supervivencia es aproximadamente del 95% mientras que en etapas mas avanzadas como la etapa IV, la cifra disminuye a menos del 25%. Sin embargo, en México, solo el 10% de los casos se detecta en etapas tempranas iniciales, mientras que el 75% de los casos se detecta en fases avanzadas.

Según datos del Instituto Nacional de Estadística y Geografía (INEGI), en 2023 se registraron 23,790 nuevos casos de cáncer de mama en personas mayores de 20 años, con una tasa de incidencia de 27.64 casos por cada 100,000 habitantes en este grupo de edad. En cuanto a la mortalidad, en 2023 se reportaron 8,034 defunciones por esta enfermedad, de las cuales el 99.5% correspondió a mujeres.

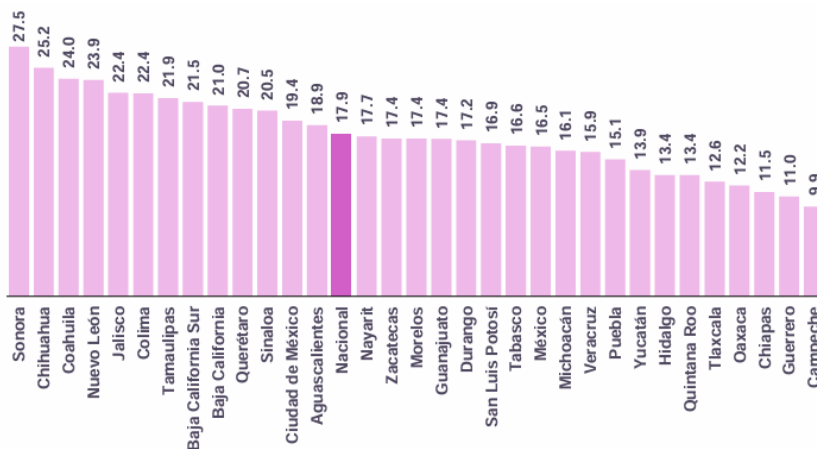


Figura 1.1: Defunciones en mujeres de 20 años y más por cáncer de mama, según entidad federativa durante el 2023. Numero de muertes por cada 100 mil mujeres [1]

La detección temprana es esencial para reducir la mortalidad por cáncer de mama. Según INEGI, el 65% de las mujeres de 50 a 59 años reportaron haberse realizado una auto exploración mamaria en 2021, y el 51.5% se sometió a una mamografía. Estas prácticas permiten identificar la enfermedad en etapas iniciales, facilitando tratamientos más efectivos y aumentando las pro-

babilidades de recuperación. Sin embargo, la cobertura de mamografías sigue siendo insuficiente, especialmente en zonas rurales y comunidades marginadas. Estos datos reflejan la necesidad de fortalecer el estudio de la detección.

Ahora bien el análisis de imágenes mediante redes neuronales es reciente. En este trabajo, se hace uso de una base de datos que es procesada con el fin de analizar si las redes neuronales convolucionales son capaces de reconocer características de anomalías en mamografías con poca calidad. Diversos estudios han demostrado que la calidad de imagen afecta de manera significativa la capacidad de los modelos de inteligencia artificial para detectar lesiones mamarias. En trabajos previos [27], se destaca que imágenes con bajo contraste, desenfoque o artefactos producen una disminución notable en la sensibilidad de modelos.

En el paper High-Resolution Breast Cancer Screening with Multi-View Deep Convolutional Neural Networks [26] se investigaron el efecto de la resolución en el rendimiento de redes profundas aplicadas a mamografías, y concluyeron que si bien los modelos pueden aprender de imágenes a resoluciones más bajas, existe un umbral mínimo de calidad por debajo del cual se pierde información diagnóstica crítica. Estos hallazgos plantean un dilema: ¿es posible diseñar o adaptar redes neuronales que conserven un rendimiento aceptable ante degradaciones inevitables?

THE EFFECT OF DECREASING THE RESOLUTION OF THE IMAGE.

scale	×1/8	×1/4	×1/2	×1
0 vs. others	0.587	0.585	0.611	0.618
1 vs. others	0.718	0.742	0.779	0.794
2 vs. others	0.729	0.750	0.777	0.787
macAUC	0.678	0.692	0.722	0.733
HC-macAUC	0.743	0.753	0.782	0.787

Figura 1.2: Efecto de la disminución de resolución en la clasificación de mamografías. AUC obtenido para diferentes escalas de resolución (de 1/8 a escala completa) utilizando un modelo de red neuronal convolucional. [26]

Otros trabajos, como en el artículo Connected-UNets: A deep learning architecture for breast mass segmentation" publicado en el periódico científico "npj Breast Cancer", donde se hace uso de la arquitectura Connected-UNets y que es evaluada en tres conjuntos de datos de mamografías: CBIS-DDSM, INbreast y un conjunto privado. Ellos delimitan la zona de interés con un ROI de 256 x 256 pixeles, logrando resultados sobresalientes. [28]

A partir de esta revisión, se identifica una brecha específica: si bien existe evidencia de que la calidad de imagen afecta el desempeño de modelos en mamografías, pocos trabajos han explorado si arquitecturas ampliamente utilizadas como ResNet50 son capaces de aprender características directamente desde imágenes de baja resolución.

En este contexto, el presente estudio se propone evaluar el rendimiento de estas arquitecturas al ser entrenadas con mamografías la máxima resolución de la base de datos y otras reducidas a una resolución menor. Esto permite observar hasta qué punto los modelos mantienen su capacidad de aprendizaje y clasificación bajo condiciones subóptimas, aportando evidencia empírica relevante para aplicaciones reales, especialmente en contextos de recursos limitados donde las imágenes de alta calidad pueden no estar disponibles.

Los resultados se analizan con la métrica IoU, métrica que usa datos que pueden significar un subajuste durante el entrenamiento.

El trabajo esta estructurado de la siguiente manera: En el primer capítulo se aborda una introducción al diagnóstico por imágenes, incluyendo un repaso de la anatomía mamaria y del cáncer de mama, desde una perspectiva científica y con fundamentos en la Física. El segundo capítulo presenta una descripción general de las redes neuronales y la segmentación semántica. En el tercer capítulo, se justifica el desarrollo del presente estudio a través de una revisión del

estado del arte. Los capítulos cuarto y quinto están dedicados a la metodología empleada y a la presentación de los resultados obtenidos, respectivamente.

Capítulo 2

Física y Medicina

2.1. Anatomía

La mama es un órgano complejo y dinámico que juega un papel fundamental en la lactancia y tiene una estructura anatómica que incluye tejido glandular y sistema de flujo linfático como las principales.

2.1.1. Sistema de flujo linfático

Los nódulos o ganglios linfáticos axilares drenan la mayor parte de la mama (75 % de la linfa) de la extremidad superior, de ahí su extrema importancia en la estadificación del cáncer de mama. Estos nódulos se dividen clásicamente en 3 niveles. Los nódulos linfáticos paraesternales drenan (25 % de la linfa) la linfa que procede de la región medial de la mama que se encuentran en el territorio de las venas torácicas internas (Figure 2.1). [5]

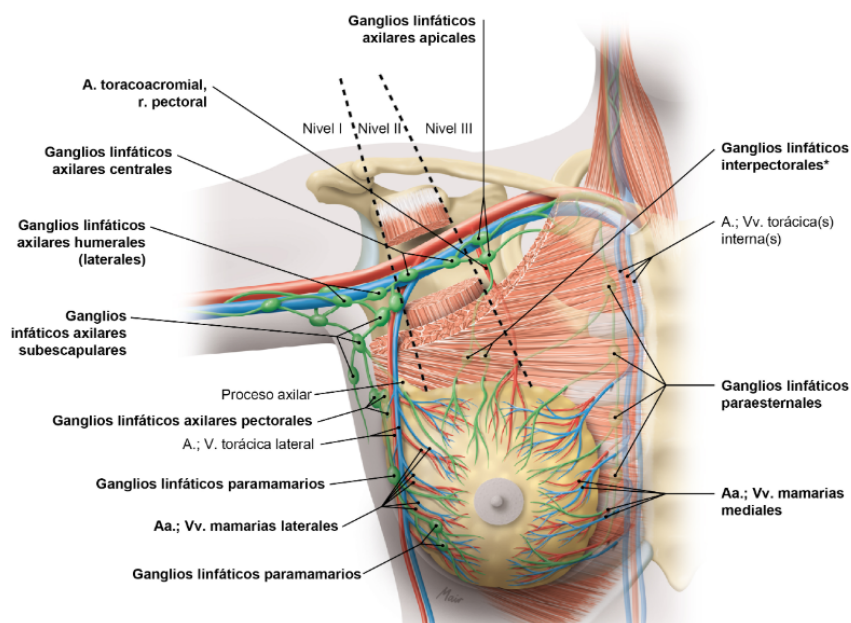


Figura 2.1: Irrigación y drenaje linfático de la mama. [5]

2.1.2. Estructura de la mama y las glándulas mamarias

1. Estructuras glandulares. presenta una prolongación hacia la axila. Está compuesta por 15-24 lóbulos glandulares, cada uno con lobulillos y un conducto galactóforo que desemboca en el pezón. Los ligamentos de Cooper la conectan a la fascia pectoral. En mujeres no lactantes, el tejido glandular está menos desarrollado. El pezón masculino se ubica en el cuarto espacio intercostal (EIC), mientras que en las mujeres su posición varía. La aréola contiene glándulas sebáceas (glándulas areolares) que facilitan la lactancia. La irrigación proviene de las arterias mamarias laterales y mediales, y la inervación corresponde a las ramas sensitivas de T2-T5.
2. Estructuras vasculares
 - Arterias: Las arterias torácicas internas irrigan la zona medial de la mama, las arterias mediales contribuyen a la irrigación en la zona lateral y las arterias torácicas laterales con arterias toracoacromiales irrigan la posición superolateral (figure 2.1).
3. Inervación
 - Nervios intercostales: Los ramos cutáneos anteriores y laterales de los nervios intercostales cuarto a sexto (T4-T6) proporcionan la sensibilidad de la piel de la mama.
4. Músculos y tejidos.
 - Musculo pectoral mayor.
 - Fascia pectoral profundo.

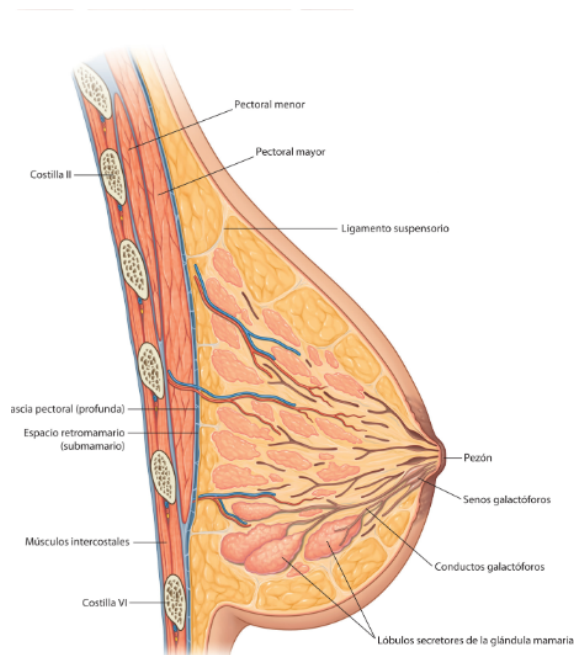


Figura 2.2: Anatomía de la Mama [6]

2.1.3. Patrones mamográficos

La densidad mamaria, clasificada por el sistema BI-RADS[®], se divide en cuatro patrones que reflejan la proporción de tejido fibroglandular respecto al tejido graso en las mamas. A medida

que aumenta la densidad, la cantidad de tejido fibroglandular es mayor, esto tiene implicaciones clínicas importantes. Las mamas con baja densidad, como en los patrones A y B, presentan una gran cantidad de tejido graso, lo que facilita la detección de lesiones en la mamografía, ya que las estructuras anormales se distinguen con mayor claridad. En cambio, las mamas heterogéneamente densas (patrón C) y extremadamente densas (patrón D) contienen más tejido fibroglandular, lo cual puede dificultar la visualización de tumores o calcificaciones, disminuyendo la sensibilidad del estudio. Además, estos patrones más densos están asociados con un aumento en el riesgo de desarrollar cáncer de mama. Por lo tanto, la mayor densidad, representa un desafío diagnóstico y un factor de riesgo adicional que puede requerir estudios complementarios.

- Predominantemente graso.
- Fibroglandular disperso o densidad media.
- Heterogéneamente denso.
- Extremadamente denso.

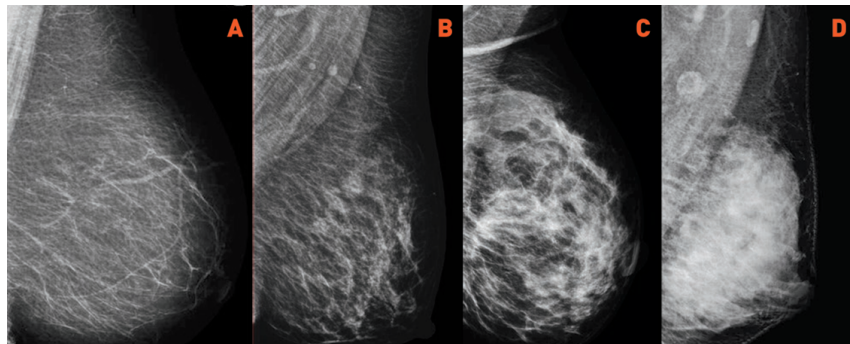


Figura 2.3: Patrones mamográficos de acuerdo con BI-RADS® 2013 [7]

2.2. Cáncer en la Mama

El cáncer de mama es una de las enfermedades más prevalentes en el mundo y una de las principales causas de mortalidad entre las mujeres. Según la Organización Mundial de la Salud (OMS), este tipo de cáncer representa el 16% de todos los cánceres femeninos y su incidencia ha aumentado en las últimas décadas. Sin embargo, gracias a los avances en la detección temprana y el tratamiento, las tasas de supervivencia han mejorado significativamente. Aproximadamente el 80% de los casos de cáncer de mama se originan en los conductos y pueden estar asociados con depósitos de calcificaciones, que se visualizan como pequeños granos de tamaño variable. Las microcalcificaciones que son calcificaciones más pequeñas, tienen un tamaño de alrededor de 500 μ m, son de particular interés. La incidencia de cáncer de mama es más frecuente en el cuadrante lateral superior de la mama (Figura 2.3). [11]

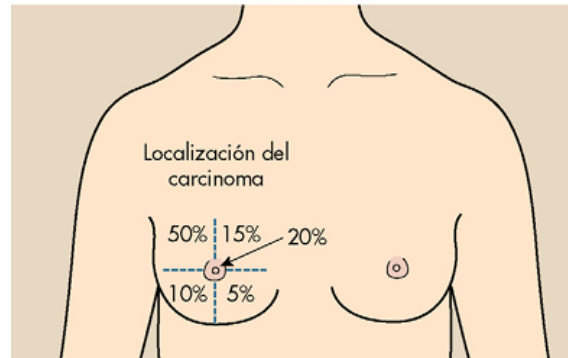


Figura 2.4: Porcentaje de incidencia de cáncer de mama según la localización en la mama. [11]

2.2.1. Clasificación radiológica

1. Nódulos; Las masas se identifican como lesiones focales con forma definida y se evalúan principalmente por su forma (ovalada, redonda, irregular), márgenes (circunscritos, microlobulados, espiculados) y densidad (alta, intermedia, baja). Las masas benignas suelen ser ovaladas y de bordes lisos, mientras que las malignas tienden a tener márgenes irregulares o espiculados y densidad alta.

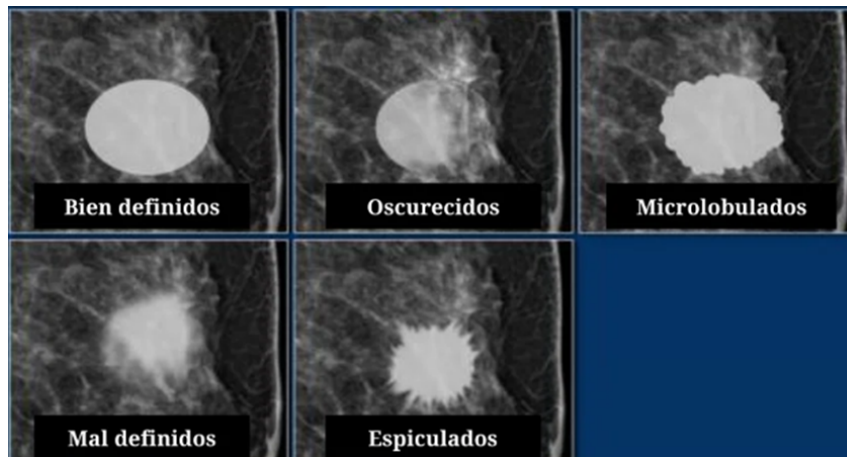


Figura 2.5: Tipos de nódulos que pueden presentarse en la mama [10]

2. Calcificaciones

Las calcificaciones son pequeños depósitos de calcio que aparecen como áreas blancas en la mamografía. Pueden ser benignas, como las calcificaciones vasculares o en palomitas de maíz, o sospechosas si son pleomórficas, finamente ramificadas o agrupadas con una distribución lineal.

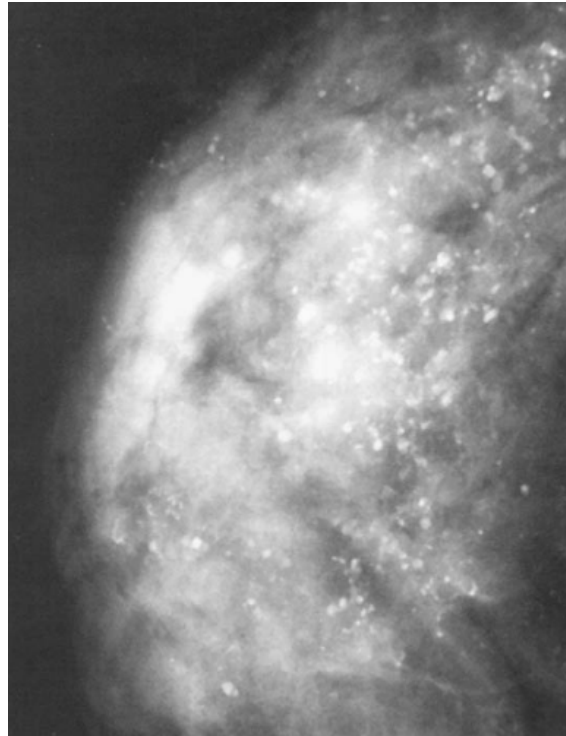


Figura 2.6: Calcificaciones [10]

3. Distorsión arquitectónica

Este hallazgo se refiere a una alteración del patrón normal del tejido mamario sin una masa definida.

4. Asimetría

Las asimetrías se presentan cuando una región de una mama tiene mayor densidad en comparación con la misma zona en la otra mama. Pueden clasificarse como asimetría focal, asimetría global, asimetría en desarrollo o asimetría verdadera. Las más preocupantes son las que aparecen nuevas en comparación con estudios previos, ya que pueden indicar una lesión en evolución. [10]

2.3. Sistema de imagen mamográfico

La física ha desempeñado un papel esencial en el desarrollo de la medicina moderna, proporcionando herramientas y métodos que han revolucionado el diagnóstico, el tratamiento y la investigación en salud. Figuras pioneras como Wilhelm Röntgen, quien descubrió los rayos X en 1895, y Marie Curie, cuyos estudios sobre la radiactividad sentaron las bases para la radioterapia, fueron fundamentales en la aplicación de principios físicos a la medicina. Asimismo, investigadores como Paul Lauterbur y Peter Mansfield, responsables del desarrollo de la resonancia magnética, han permitido una comprensión más profunda del cuerpo humano y la mejora de diversas técnicas médicas.

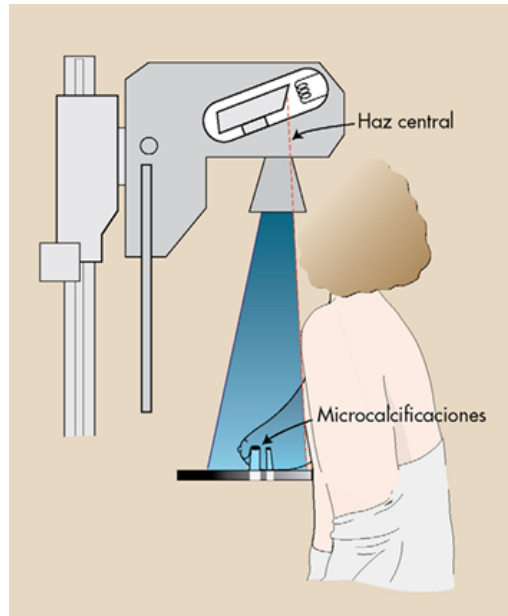


Figura 2.7: Sistema de mamografía. [11]

En esta sección hablaremos de los principios físicos de la mamografía, técnica que permiten a los profesionales de la salud visualizar de estructura de la mama de manera no invasiva, facilitando el diagnóstico temprano y preciso de diversas enfermedades.

La mamografía utiliza rayos X de baja energía, también conocidos como radiación de espectro blando. Estos rayos X tienen longitudes de onda más largas y menor penetración en comparación con los rayos X utilizados en otras modalidades de diagnóstico, como la tomografía computarizada. Esto es ideal para visualizar el tejido mamario, que tiene una densidad similar a la de los tejidos blandos.

2.3.1. Generador de alta frecuencia

El propósito de un generador de alta frecuencia en los sistemas de mamografía es convertir la corriente alterna estándar de la red eléctrica en una corriente continua de alto voltaje con mínima fluctuación.

El principio físico comienza cuando el generador toma la corriente alterna monofásica y la rectifica, convirtiéndola en corriente continua. Esta corriente continua alimenta un circuito inversor que la fragmenta a una frecuencia muy alta, típicamente entre 5 y 10 kHz. Este paso es crucial porque aumentar la frecuencia permite el uso de un transformador de alto voltaje mucho más pequeño y eficiente para elevar el voltaje al nivel requerido para la producción de rayos X (generalmente entre 19 y 35 kilovolts en mamografía).

Una vez que el voltaje ha sido elevado por este transformador de alta frecuencia, la señal es rectificadas nuevamente y filtrada a través de condensadores. Este proceso final de suavizado elimina las fluctuaciones, produciendo una salida de voltaje de corriente continua casi perfecta, con una variación inferior al 1%. Este voltaje constante asegura que los electrones acelerados desde el cátodo hacia el ánodo en el tubo de rayos X lo hagan con una energía cinética constante y predecible. Esto resulta en la producción de un haz de fotones de rayos X más homogéneo y eficiente, lo que es fundamental para obtener imágenes de alta calidad con la menor dosis de radiación posible.

2.3.2. Tubo de rayos X

El tubo de rayos X es el componente central de los equipos de mamografía. Está compuesto por un cátodo y un ánodo encerrados en un vacío parcial.

El cátodo, generalmente hecho de filamento de tungsteno, emite electrones cuando se calienta mediante una corriente eléctrica. El tungsteno tiene un punto de fusión de aproximadamente 3400° . Debido a una diferencia de voltaje de decenas a cientos de Kilovolts los electrones del cátodo son acelerados hacia el ánodo, donde se encuentra el blanco, que para sistemas de mamografía está hecho de molibdeno y rodio por su número atómico bajo y la baja energía de su capa electrónica K. [12]

Elemento	Símbolo	Numero atómico	Energía de los rayos X de la capa K	Temperatura de fusión
Tungsteno	W	74	69keV	3400°
Molibdeno	Mo	42	19keV	2600°
Rodio	Rh	45	23keV	3200 °

Tabla 2.1: Rayos X resultantes de la capa K. [12]

Cuando los electrones impactan el ánodo, su energía cinética se convierte en rayos X mediante dos procesos principales: la radiación de frenado (bremsstrahlung) y la emisión característica. [13]

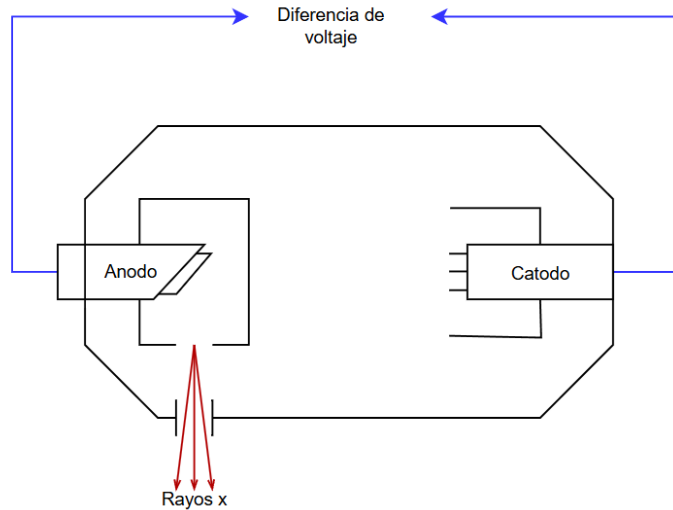


Figura 2.8: Tubo de rayos x

Radicación característica

La radiación característica, como la producida por el tungsteno, molibdeno y rodio, es un tipo de emisión de rayos X que surge cuando electrones acelerados impactan un átomo blanco y desplazan electrones internos de sus capas atómicas profundas como la K o L. Este desplazamiento crea huecos que son ocupados por electrones de capas superiores, liberando fotones con energías discretas específicas del material. En el caso del molibdeno, con número atómico 42, las transiciones electrónicas más importantes generan líneas espectrales características como $k\alpha$ (aproximadamente 59.3 keV) cuando un electrón de la capa L llena un vacío en la capa K, y $K\beta$ (alrededor de 67.2 keV) cuando M llena un vacío en la capa K. Estas emisiones forman picos agudos en el espectro de rayos X, contrastando con el espectro continuo del bremsstrahlung [14].

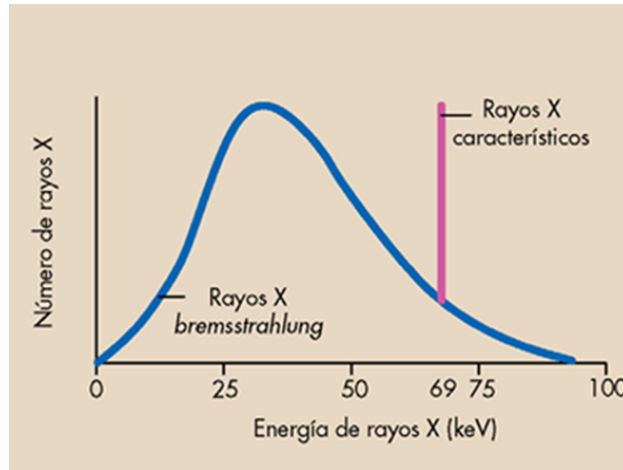


Figura 2.9: Espectro común de la radiación de frenado y radiación característica.

El proceso requiere que la energía del electrón incidente supere la energía de enlace de los electrones atómicos internos, siendo para la capa K del tungsteno de 69.5 keV. La radiación característica es fundamental en aplicaciones como los tubos de rayos X médicos, donde el tungsteno se emplea como ánodo por su alta eficiencia en producir estos rayos X característicos, y en técnicas de difracción de rayos X para análisis de materiales. A diferencia de la radiación de frenado, que produce un espectro continuo al desacelerar electrones, la radiación característica muestra líneas espectrales discretas con energías bien definidas determinadas por la estructura electrónica del átomo blanco. [15]

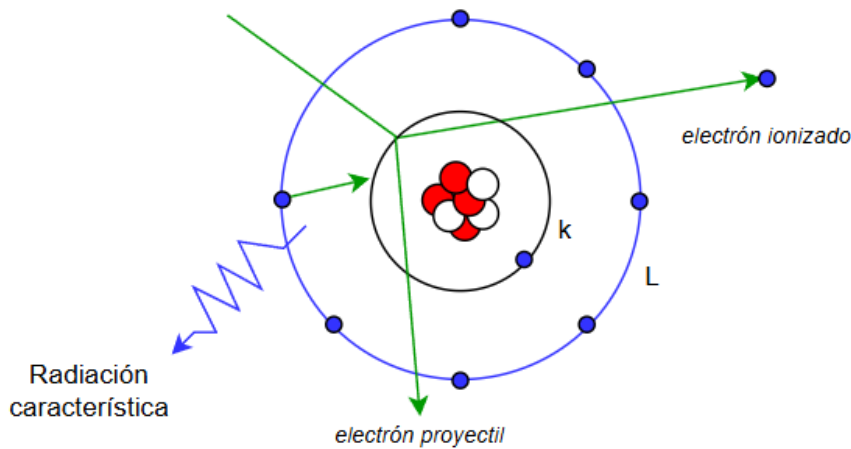


Figura 2.10: Mecanismo para la producción de rayos X característicos.

Radiación de frenado

Cuando un electrón es acelerado por una diferencia de potencial ΔV , adquiere una energía cinética $K = e\Delta V$. Al chocar con los átomos de un material blanco, el electrón se desacelera y pierde parte de esa energía, la cual se emite en forma de radiación electromagnética (fotones), fenómeno conocido como Bremsstrahlung. Si el electrón tenía una energía cinética K antes del choque y K' después, la energía del fotón emitido será $hf = K - K' = \frac{hc}{\lambda}$. Dado que los electrones

pueden tener múltiples colisiones, se genera un espectro continuo de rayos X con fotones de distintas energías. La energía máxima de un fotón ocurre cuando el electrón pierde toda su energía en una sola colisión ($K' = 0$), y en ese caso se obtiene la longitud de onda mínima posible $\lambda_{\text{mín}} = \frac{hc}{K} = \frac{hc}{e\Delta V}$, que representa el límite inferior del espectro emitido. [16]

La radiación de frenado se produce cuando electrones acelerados por voltajes típicos de 10,000 V chocan con un material, frenan y emiten fotones en el rango de los rayos X, con longitudes de onda mínimas de unas décimas de nanómetro. Este proceso genera un espectro continuo de radiación, distinto de las líneas discretas de los rayos X producidos en transiciones atómicas. La interacción se representa simbólicamente como la emisión de un fotón por un electrón en desaceleración, proceso inverso al efecto fotoeléctrico. Ambos fenómenos requieren la presencia de un átomo pesado para conservar el momento, ya que no pueden ocurrir con electrones libres. [16]

En palabras sencillas la radiación de frenado es producido por el cambio de la energía cinética del electrón proyectil a energía electromagnética. cuando el electrón proyectil pasa cerca del núcleo desacelera quedando con menor energía cinética. La pérdida de energía cinética reaparece en forma de rayos X. [13]

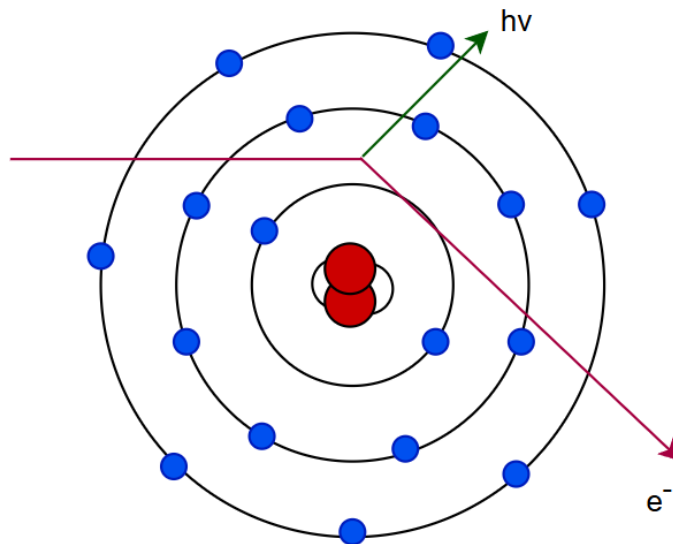


Figura 2.11: Radiación de frenado.

2.3.3. Diana del tubo de rayos x

Los tubos de rayos X en mamografía utilizan dianas de tungsteno (W), molibdeno (Mo) o rodio (Rh). Los rayos X más útiles para maximizar el contraste en la mama tienen energías entre 17 y 24 keV.

El tungsteno emite principalmente radiación de frenado (bremsstrahlung), con rayos X característicos de 12 keV que no contribuyen a la imagen, solo a la dosis del paciente. El molibdeno genera principalmente rayos X característicos de 17 y 19 keV, ideales para mamografía, con poca radiación de frenado debido a su menor número atómico ($Z = 42$). El rodio, con $Z = 45$, produce un espectro similar pero con un pico característico a 23 keV y más radiación de frenado.

Los sistemas modernos de mamografía combinan diana y filtro Mo/Mo, Mo/Rh o Rh/Rh, optimizando la calidad de imagen y reduciendo la dosis de radiación.

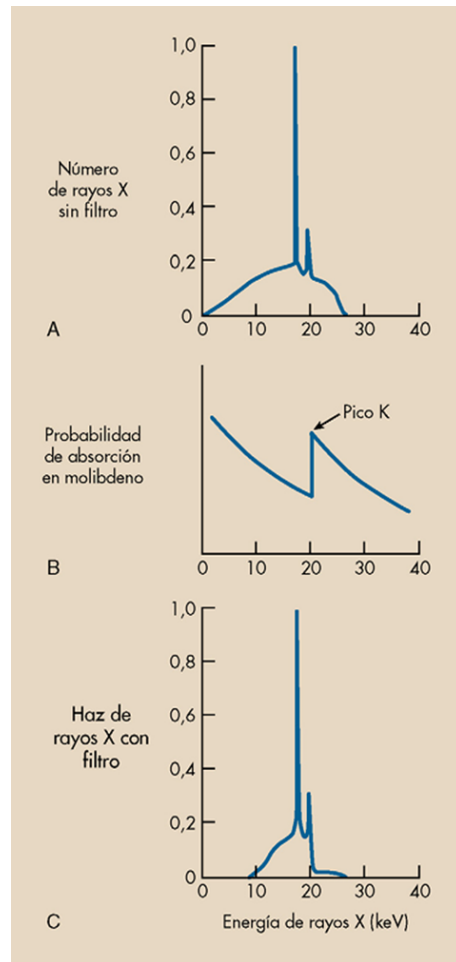


Figura 2.12: Espectros de emisión de una diana de Molibdeno a) sin filtro c) con filtro de molibdeno. b) Probabilidad de absorción

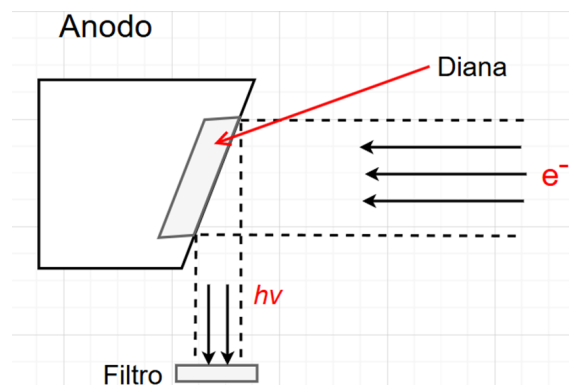


Figura 2.13: Diana y filtro de un tubo de rayos X

2.3.4. Punto focal

El tamaño del punto focal es crucial en los tubos de rayos X para mamografía debido a la alta demanda de resolución espacial necesaria para detectar calcificaciones pequeñas (microcalcificaciones). Los tubos de mamografía suelen utilizar puntos focales de entre 0,3 y 0,1 mm de diámetro, siendo preferibles los más pequeños. Además del tamaño, la forma del punto focal también es importante; los puntos focales circulares son ideales, aunque los rectangulares son comunes. Los fabricantes controlan la forma y el tamaño del punto focal mediante el diseño del cátodo y una estructura llamada copa focalizadora. Sin embargo, existe una gran variabilidad en el tamaño del punto focal declarado, por lo que es esencial que los físicos médicos realicen pruebas de aceptación para verificar su tamaño y la resolución espacial.

Para lograr un punto focal pequeño y una intensidad de rayos X adecuada, los fabricantes utilizan el principio de enfoque en línea y la inclinación del tubo de rayos X. Un punto focal eficaz de 0,3/0,1 mm se obtiene con un ángulo del ánodo de aproximadamente 23 grados y una inclinación del tubo de 6 grados. El cátodo se coloca hacia la pared torácica, lo que facilita la ubicación del paciente y aprovecha el efecto talón del ánodo. Esta inclinación también permite capturar imágenes del tejido cercano a la pared torácica, asegurando que no se pierda tejido durante la toma de imágenes.

2.3.5. Interacción de los Rayos X con la Mama

Efecto fotoeléctrico

El efecto fotoeléctrico ocurre cuando un fotón de rayos X transfiere toda su energía a un átomo, expulsando un fotoelectrón de una capa interna. Un electrón externo ocupa su lugar, liberando un nuevo fotón de menor energía. Entonces, Si la energía del fotón $h\nu$ supera la función de trabajo, la energía excedente aparece como la energía cinética del electrón. [17]

$$E_c = h\nu - \phi \quad (2.1)$$

Donde:

- E_c es la energía cinética
- h es la constante de Planck,
- ν es la frecuencia del fotón,
- ϕ Función de trabajo del material.

Aumentar la intensidad solo incrementa el número de fotoelectrones, no su energía. El efecto es más fuerte cuando la energía del fotón es apenas mayor que la de ligadura de un electrón interno, siendo relevante en la capa K. En tejidos blandos, la atenuación fotoeléctrica disminuye a partir de 35 keV, pero sigue siendo importante en huesos hasta 50 keV. [17]

El efecto fotoeléctrico tiene mayor probabilidad de ocurrir cuando el fotón interactúa con elementos de número atómico elevado y a energías relativamente bajas (entre 10–30 keV).

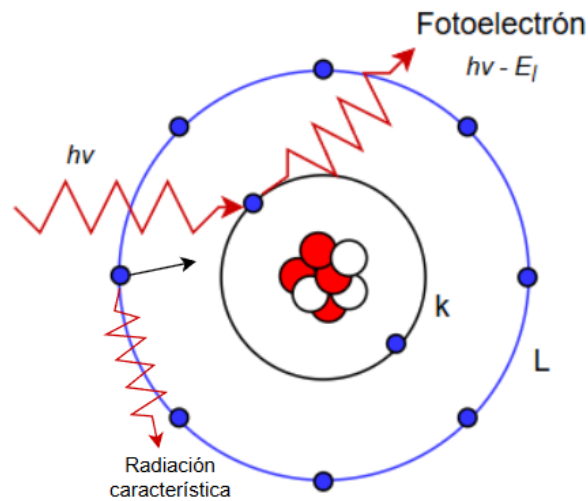


Figura 2.14: Efecto fotoeléctrico

Efecto Compton

El efecto Compton es el fenómeno donde un fotón de alta energía (como los rayos X o gamma) choca con un electrón libre o débilmente ligado, transfiriéndole parte de su energía y momento, lo que causa un aumento en la longitud de onda del fotón dispersado. Fue descubierto por Arthur Compton en 1923 y confirma la naturaleza cuántica de la luz. [18]

Matemáticamente, el cambio en la longitud de onda ($\Delta\lambda$) del fotón dispersado depende del ángulo de dispersión (θ) y se describe como:

$$\Delta\lambda = \lambda_f - \lambda_i = \frac{h}{m_e c} (1 - \cos \theta),$$

donde:

- λ_i y λ_f son las longitudes de onda inicial y final del fotón.
- h es la constante de Planck.
- m_e es la masa del electrón.
- c es la velocidad de la luz.
- θ es el ángulo de dispersión del fotón.

El término $\frac{h}{m_e c} \approx 2,43 \times 10^{-12}$ m, se conoce como la longitud de onda Compton del electrón. La energía del fotón dispersado (E_f) se calcula como:

$$E_f = \frac{E_i}{1 + \frac{E_i}{m_e c^2} (1 - \cos \theta)},$$

donde E_i es la energía inicial del fotón. [19]

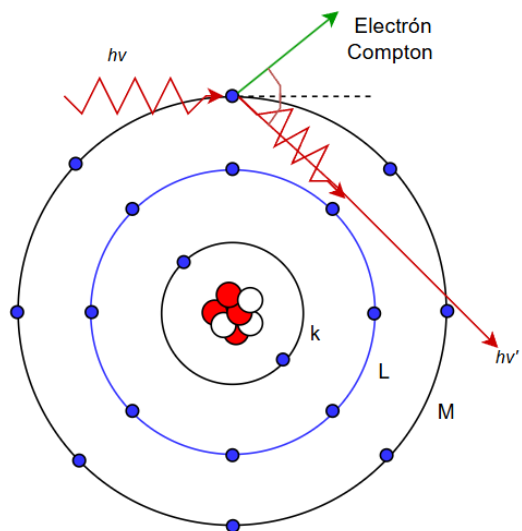


Figura 2.15: Efecto Compton

2.3.6. Captura de la imagen

En mamografía se han empleado cuatro tipos de receptores de imagen, pero actualmente solo se usan pantalla-película y detectores digitales. La pantalla-película utiliza películas de emulsión por una cara combinadas con una pantalla negra para evitar el cruzamiento de luz, lo que mejora el contraste. Estas se colocan en cassetes especiales con cubiertas de baja absorción y un cierre diseñado para un buen contacto.

El uso de pantallas intensificadoras aumenta la velocidad del sistema, reduciendo la dosis a la paciente y mejorando el contraste radiográfico. Para una mejor resolución espacial, la película debe colocarse entre el tubo de rayos X y la pantalla, con la emulsión orientada hacia esta última.

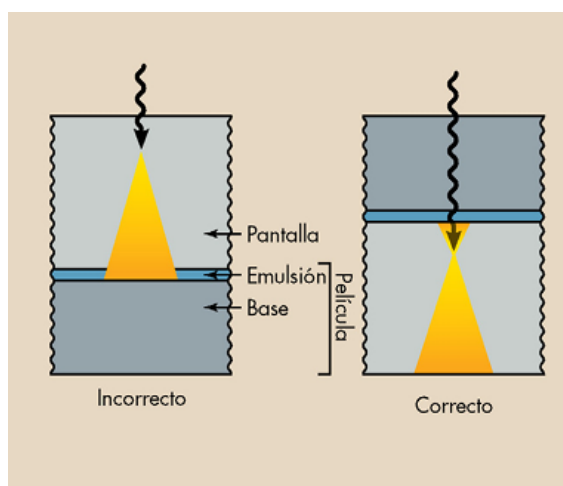


Figura 2.16: Manera correcta de colocar la película mamográfica y situar la casete. [11]

Capítulo 3

Estado del arte

3.1. Introducción

Las redes neuronales artificiales (RNA) son modelos de aprendizaje automático que tienen la intención de imitar el comportamiento de las neuronas biológicas. La unidad más pequeña de una red neuronal artificial es la neurona o también llamada capa. Simula varias funciones básicas de la neurona biológica, evalúa la intensidad de cada entrada, suma las diferentes entradas y compara el resultado con un umbral apropiado. Por último, determina cuál es el valor de salida. [20]. Reciben el nombre de capa de entrada, capas ocultas y la capa de salida respectivamente.

Cada neurona realiza una operación sencilla que consiste en activarse si la cantidad total de señal recibida supera un umbral de activación. Si una neurona se activa, emite una señal que se transmite a otras neuronas a las que está conectada. Cada nodo o punto de conexión actúa como un filtro que transforma el mensaje recibido en una señal. El enlace entrada-salida no está programado, sino que simplemente se obtiene a partir de un proceso de aprendizaje basado en datos empíricos que pueden ser de este tipo: aprendizaje supervisado y no supervisado. [21]

En el **aprendizaje supervisado** se le proporciona al modelo un conjunto de datos en el que cada entrada (vectores) está emparejada con la salida correcta. El modelo aprende a mapear las entradas con las salidas correctas ajustando sus parámetros internos para minimizar el error entre las salidas predichas y las salidas reales. Esto se hace a través de un proceso iterativo en el que el modelo se entrena utilizando el conjunto de datos etiquetado.

Por otra parte, en el **aprendizaje no supervisado**, el modelo recibe solo las entradas y debe encontrar relaciones, agrupaciones o estructuras dentro de los datos. Esto puede incluir tareas como la agrupación (clustering), donde el modelo agrupa datos similares, o la reducción de dimensionalidad, donde el modelo encuentra formas más compactas de representar los datos. [22]

3.2. Neuronas

En 1943 Warren S. McCulloch y Walter Pitts publicaron "A logical calculus of the ideas immanent in nervous activity" en este artículo se intenta comprender cómo el cerebro humano podría producir patrones complejos a través de las neuronas. Una de las principales ideas que surgieron en su trabajo fue la comparación de neuronas con un umbral binario con la lógica booleana. Su modelo utiliza una función de activación basada en un umbral. Si la suma ponderada de las entradas supera o iguala este umbral, la neurona se activa (1), si no, permanece inactiva (0).

Sea x_1, x_2, \dots, x_n las entradas, w_1, w_2, \dots, w_n los pesos asociados y b conocido como bias (sesgo), que ajusta el hiperplano de decisión.

$$z = b + \sum_{i=1}^n x_i w_i \quad (3.1)$$

Entonces, la salida (y) se define como:

$$y = \begin{cases} 1, & z \geq \theta \\ 0, & z < \theta \end{cases} \quad (3.2)$$

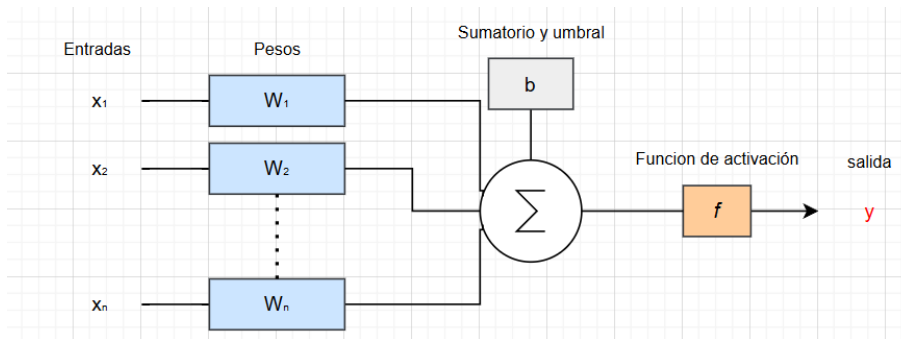


Figura 3.1: Arquitectura de una neurona artificial

Posteriormente en el año de 1962 Hubel y Wiesel realizaron estudios sobre la corteza visual primaria en gatos y monos revelando que ciertas neuronas responden específicamente a estímulos visuales simples, como bordes, líneas y movimientos en regiones delimitadas del campo visual, denominadas campos receptivos. Identificaron dos tipos principales de neuronas:

1. Neuronas simples, que detectan líneas y bordes con orientaciones específicas en posiciones concretas del campo visual.
2. Neuronas complejas, que responden a estímulos similares, pero con mayor invariancia a su posición y con sensibilidad al movimiento.

Estos hallazgos evidenciaron una organización jerárquica en el sistema visual, donde las características simples se combinan para formar representaciones más complejas en niveles superiores

3.3. El Modelo multicapa

El Modelo multicapa en redes neuronales, desarrollado por Frank Rosenblatt en 1958 amplía la capacidad de los modelos de una sola capa al incorporar una o más capas ocultas entre la capa de entrada y la de salida. Cada capa está compuesta por múltiples neuronas interconectadas, donde cada neurona de una capa está conectada a todas las neuronas de la siguiente capa.

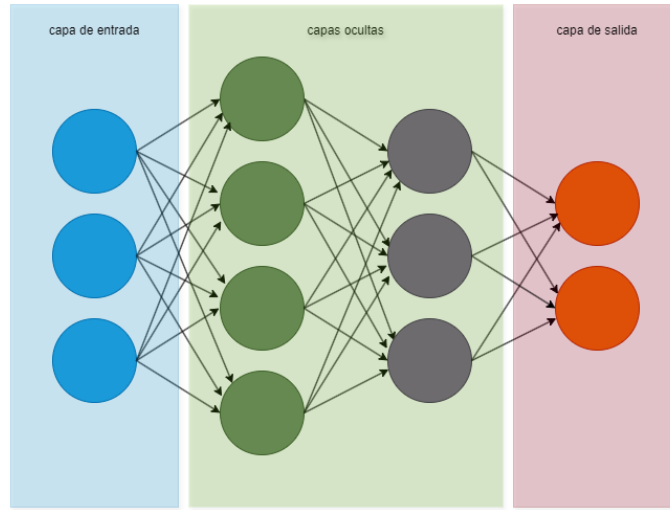


Figura 3.2: Ejemplo de la estructura de una red neuronal artificial (modelo multicapa)

En cada neurona, se calcula una combinación lineal de las entradas ponderadas y se añade un término de sesgo. A este resultado se le aplica una función de activación no lineal, como la sigmoide o la tangente hiperbólica, para producir la salida de la neurona. Este proceso se repite capa por capa, propagando la información desde la entrada hasta la salida.

El entrenamiento de un modelo multicapa se realiza utilizando el algoritmo de retropropagación del error. Este método ajusta los pesos y sesgos de la red para minimizar la diferencia entre las salidas predichas y los valores reales del conjunto de entrenamiento. Este proceso implica calcular el error en la capa de salida y luego retropropagar este error hacia atrás a través de la red, actualizando los pesos y sesgos en función de las derivadas parciales del error con respecto a cada parámetro.

De la ecuación 3.1 se puede obtener que para una neurona j en la capa l , la entrada neta z_j^l se calcula como:

$$z_j^l = \sum_i w_{ji}^l a_i^{l-1} + b_j^l \quad (3.3)$$

donde:

- w_{ji}^l es el peso que conecta la neurona i de la capa $l - 1$ con la neurona j de la capa l ,
- a_i^{l-1} es la activación de la neurona i en la capa anterior,
- b_j^l es el sesgo de la neurona j en la capa l .

La activación a_j^l de la neurona j en la capa l se obtiene aplicando la función de activación σ :

$$a_j^l = \sigma(z_j^l) \quad (3.4)$$

El error δ_j^l de la neurona j en la capa l se define como:

$$\delta_j^l = \frac{\partial C}{\partial z_j^l} \quad (3.5)$$

donde C es la función de coste que mide la discrepancia entre la salida deseada y la salida real de la red.

Las derivadas parciales de C con respecto a los pesos y sesgos se calculan como:

$$\frac{\partial C}{\partial w_{ji}^l} = a_i^{l-1} \delta_j^l \quad (3.6)$$

$$\frac{\partial C}{\partial b_j^l} = \delta_j^l \quad (3.7)$$

Estas derivadas se utilizan para actualizar los pesos y sesgos mediante un método de optimización, como el descenso de gradiente.

La introducción de capas ocultas y el uso de funciones de activación no lineales permiten que el modelo multicapa aprenda representaciones internas complejas, lo que lo hace adecuado para muchas aplicaciones en aprendizaje automático.

3.4. Redes neuronales convolucionales

Las redes neuronales convolucionales [LeCun et al., 1989] son un tipo de red neuronal principalmente usada en la visión artificial que a diferencia de las Redes neuronales tradicionales fueron creadas con el propósito de usarse en imágenes. Gracias a sus capas de convolución que trabajan con filtros estas redes neuronales tienen una gran capacidad de reconocer diferentes características en una imagen. Convirtiéndolas en una herramienta importante para la clasificación y detección de objetos.

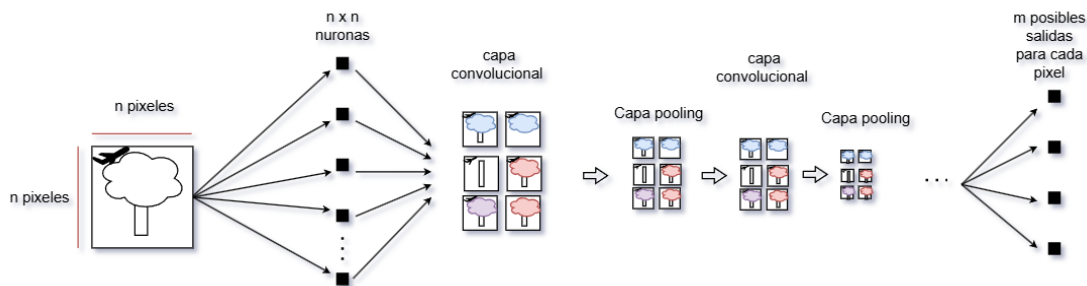


Figura 3.3: Ejemplo de la estructura de una CNN común.

Una red neuronal convolucionales, tiene estructura muy similar a las redes neuronales tradicionales, pero que implementa dos capas completamente diferentes. Veamos entonces cual es la funcionalidad de estas dos capas.

1. Capas de Convolución (Convolutional Layers):

Estas capas no solo le dan el nombre a las redes si no también son el núcleo de las CNNs. Aplican filtros (kernels) a la entrada para crear mapas de características. Los filtros son pequeñas matrices que recorren la imagen de entrada y detectan características como bordes, texturas y patrones específicos. La operación de convolución ayuda a reducir la dimensión espacial de la imagen mientras mantiene la relación espacial entre los píxeles. [24]

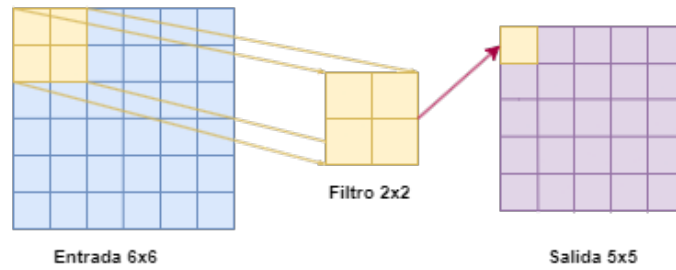


Figura 3.4: Operación de convolución [24].

Este ejemplo se puede escribir como

$$S(i, j) = (K * I)(i, j) = \sum_m \sum_n I(m, n)K(i - m, j - n) \quad (3.8)$$

Donde K es la imagen de entrada e I es el kernel o filtro, S es la imagen de salida.

2. Capas de Pooling (Pooling Layers): Estas capas reducen la dimensión de los mapas de características generados por las capas de convolución, disminuyendo la carga computacional y la posibilidad de sobre ajuste. El tipo más común es el max pooling, que toma el valor máximo en una ventana (por ejemplo, 2x2) y lo usa para crear un mapa de características más pequeño. [25]

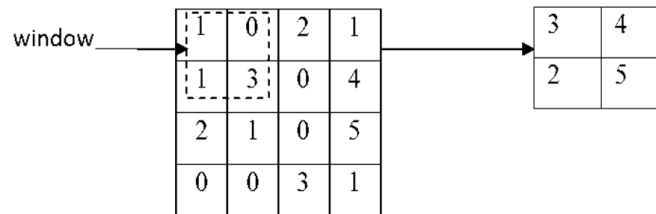


Figura 3.5: Operación de agrupación realizada eligiendo una ventana de 2 x 2. [25]

Estas dos, son capas principales e importantes en la arquitectura de las CNN, esto no quiere decir que estas capas sean las únicas presentes en una red, simplemente son las que caracterizan a las redes neuronales convolucionales.

3.4.1. Resnet y los bloque residuales

ResNet es una red neuronal convolucional (CNN) profunda desarrollada por Kaiming He y otros colegas de Microsoft Research, presentada en el paper “Deep Residual Learning for Image Recognition” en 2015 [37]. Las redes neuronales con grandes cantidades de capas sufren un problema conocido como pérdida del gradiente en arquitecturas muy profundas. Este problema ocurre cuando al aumentar la profundidad de la red, el rendimiento disminuye debido a que los gradientes se hacen muy pequeños o desaparecen al retropropagarse.

ResNet introdujo el concepto de bloques residuales para permitir un entrenamiento más efectivo en redes profundas. La clave es la inclusión de conexiones de atajo (skip connections), que permiten que la entrada de un bloque salte directamente hacia su salida. Cada bloque (Figura 2.6) puede expresarse de forma general:

$$y_i = \mathcal{F}(w_i, W_i) + x_i, \tag{3.9}$$

$$x_{i+1} = f(y_i) \tag{3.10}$$

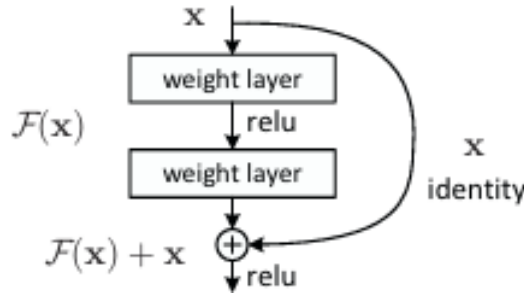


Figura 3.6: Aprendizaje residual [37]

donde: x_i es la entrada y x_{i+1} la salida al bloque residual i -ésima, $\mathcal{F}(w_i, W_i)$ es la transformación aprendida que incluye convoluciones, normalización por lotes y activaciones no lineales, x_i es un mapeo de identidad y f es la función ReLU

Esta estructura facilita que la red aprenda correcciones residuales ($\mathcal{F}(x)$) en lugar de un mapeo completo, lo que reduce la dificultad del aprendizaje y acelera la convergencia, además así evitan la pérdida del gradiente.

3.4.2. Resnet- 50

ResNet-50 es una versión popular de ResNet, que consta de 50 capas de convolución. Su arquitectura está diseñada con bloques residuales que incorporan convoluciones 1×1 , 3×3 y 1×1 , con 64, 256, 128, 512, 1024 y 2048 filtros, además, ResNet-50 utiliza los bloques residuales de tipo "bottleneck" (Figura 3.7), donde las dimensiones se reducen y se restaura dentro del bloque. Esta versión fue presentada en el mismo paper donde se presentó la arquitectura Resnet general.

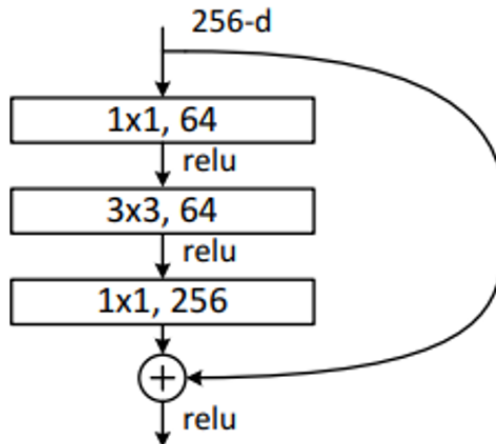


Figura 3.7: Aprendizaje residual [37]

Nombre del bloque	Tamaño de salida	ResNet-50	
conv1	112x112	7x7, 64, stride 2 3x3 max pool, stride 2	
conv2_x	56 × 56	1 × 1, 64 3 × 3, 64 1 × 1, 256	× 3
conv3_x	28x28	1 × 1, 128 3 × 3, 128 1 × 1, 512	× 4
conv4_x	14x14	1 × 1, 256 3 × 3, 256 1 × 1, 1024	× 6
conv5_x	7x7	1 × 1, 512 3 × 3, 512 1 × 1, 2048	× 3
1x1	1x1	average pool 1000-d fc, softmax	
FLOPs		3,8 × 10 ⁹	

Tabla 3.1: arquitectura de ResNet-50 [37]

3.5. Segmentación semántica

La segmentación de imágenes constituye una de las tareas fundamentales en el ámbito de la visión por computadora, con aplicaciones que abarcan desde el diagnóstico médico hasta la conducción autónoma. Su objetivo principal radica en dividir una imagen en regiones significativas, asignando una etiqueta a cada píxel o grupo de píxeles permitiendo la identificación y delimitación precisa de objetos o regiones.

3.5.1. DeepLabV3+

DeepLabV3+ es un modelo de segmentación semántica, desarrollado por Google Research en 2018. DeepLabV3+ introduce una serie de innovaciones que mejoran significativamente la precisión y eficiencia en la segmentación semántica. En primer lugar, su arquitectura se basa en un esquema encoder-decoder, donde el encoder extrae características de la imagen mediante una red convolucional profunda, como ResNet o Xception, y el decoder refina la segmentación, especialmente en los bordes de los objetos. Esta combinación permite una representación más detallada y precisa de las regiones segmentadas.

Otra característica clave es el uso de convoluciones atrous (también conocidas como convoluciones dilatadas), las cuales permiten aumentar el campo receptivo de la red sin incrementar el número de parámetros o reducir la resolución espacial. Esto es fundamental para capturar información contextual a múltiples escalas, lo que mejora la capacidad del modelo para entender el contexto global de la imagen.

Además, DeepLabV3+ incorpora el módulo Atrous Spatial Pyramid Pooling (ASPP), que aplica convoluciones atrous con diferentes tasas de dilatación en paralelo. Este enfoque multi-escala permite al modelo capturar tanto detalles locales como contextos globales, lo que resulta en una segmentación más robusta y precisa. Finalmente, el modelo fusiona características de bajo nivel (detalladas) con características de alto nivel (semánticas) en el decoder, lo que refina aún más la segmentación, especialmente en los bordes de los objetos.

DeepLabV3+ se basa en técnicas avanzadas de redes neuronales convolucionales (CNN), convoluciones atrous y el módulo ASPP. A continuación, se describen los componentes clave del modelo.

Convolución Atrous

La convolución atrous es una extensión de la convolución tradicional que introduce un parámetro de dilatación r . Mientras que la convolución tradicional se define en la ecuación 3.8, la convolución atrous se expresa como:

$$S[i, j] = \sum_{m, n} K[i - r \cdot m, j - r \cdot n] \cdot I[m, n]. \quad (3.11)$$

Aquí, r controla el espaciado entre los valores del kernel, permitiendo un campo receptivo más amplio sin aumentar el costo computacional.

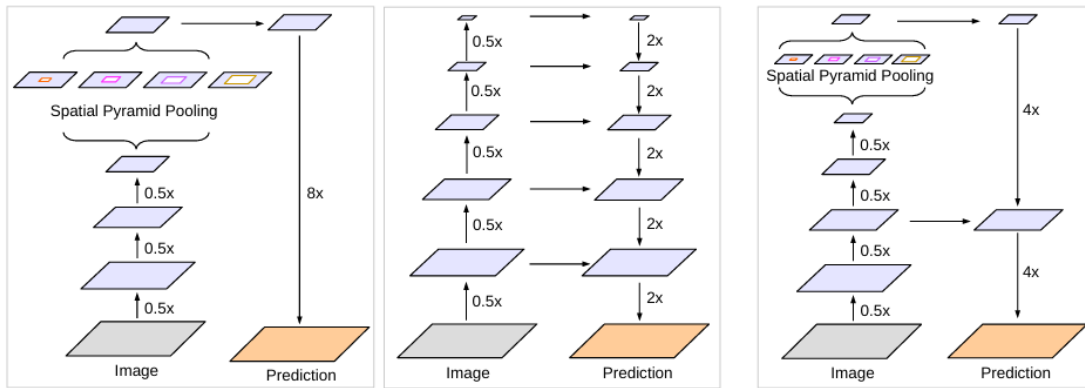


Figura 3.8: La primera y tercera arquitectura incorporan Spatial Pyramid Pooling para capturar información multiescala, mientras que la segunda sigue un enfoque clásico con múltiples niveles de resolución. Las diferencias en la escalabilidad y la propagación de características afectan la precisión de la predicción.

Atrous Spatial Pyramid Pooling (ASPP)

El módulo ASPP aplica convoluciones atrous con diferentes tasas de dilatación r_1, r_2, \dots, r_n en paralelo. Para una entrada x , cada convolución atrous produce una salida y_k :

$$y_k = \text{ConvAtrous}(x, r_k). \quad (3.12)$$

Estas salidas se combinan para formar una representación multi-escala, que incluye información contextual a diferentes niveles de resolución.

Arquitectura Encoder-Decoder

El encoder utiliza una red convolucional profunda para extraer características de la imagen, mientras que el decoder fusiona características de bajo nivel (detalladas) con características de alto nivel (semánticas). La operación de fusión en el decoder puede expresarse como:

$$y_{\text{decoder}} = f(y_{\text{encoder}}, y_{\text{low-level}}), \quad (3.13)$$

donde f es una función de fusión, como la concatenación seguida de una convolución.

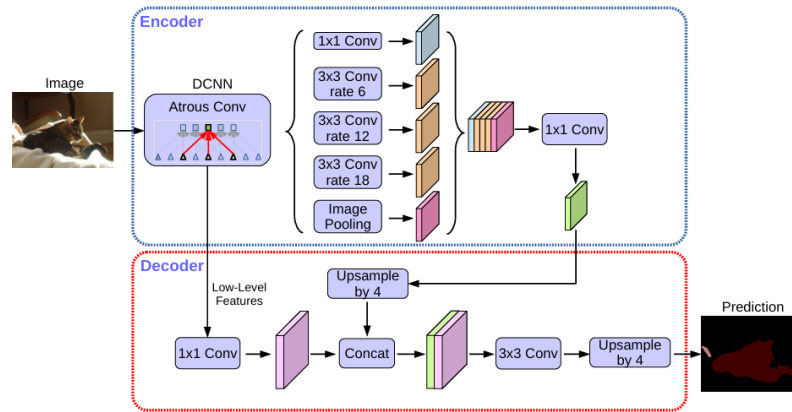


Figura 3.9: Arquitectura Ecoder-Decoder

3.6. Función de Pérdida

DeepLabV3+ utiliza una función de pérdida basada en la entropía cruzada (cross-entropy) para entrenar el modelo. La pérdida se calcula comparando la salida de la red (un mapa de probabilidades por clase) con la etiqueta verdadera (ground truth):

$$L = -\frac{1}{N} \sum_n \sum_{i,j} \sum_c y_{n,i,j,c} \cdot \log(p_{n,i,j,c}) \quad (3.14)$$

donde:

- n : índice de imagen en el batch, con N imágenes en total.
- i, j coordenadas de píxeles (altura y anchura).
- c índice de la clase.
- $y_{n,i,j,c}$ valor del ground truth codificado en one-hot para la imagen n , en el píxel (i, j) , para la clase c .
- $p_{n,i,j,c}$ probabilidad predicha por el modelo para ese mismo píxel y clase.

3.6.1. Regularización L2

La regularización L2 también conocida como regularización de peso o regularización de Tikhonov, es una técnica utilizada en el entrenamiento de redes neuronales para prevenir el sobreajuste (overfitting).

Durante el entrenamiento, el modelo aprende a minimizar una función de pérdida (por ejemplo, error cuadrático, entropía cruzada, etc.). Con la regularización L2, se agrega un término adicional a esta función de pérdida que penaliza los pesos grandes. La nueva función de pérdida se ve así:

$$L_{total} = L + \lambda \sum_{i=1}^n w_i^2 \quad (3.15)$$

donde:

λ es el coeficiente de regularización, un hiperparámetro que controla cuánto se penalizan los pesos grandes y w_i son los pesos del modelo.

Este término adicional fuerza al modelo a mantener los pesos pequeños, lo que tiende a hacer que las predicciones sean más suaves y menos sensibles a pequeñas fluctuaciones en los datos.

3.6.2. Descenso del gradiente

El SGD con momentum (SGDM) optimiza los pesos del modelo utilizando mini-batches de datos, combinando la eficiencia del entrenamiento estocástico con la estabilidad del momentum. En lugar de actualizar los pesos directamente con el gradiente actual, SGDM acumula un promedio móvil de gradientes pasados (velocidad), lo que reduce la variabilidad (ruido) de las actualizaciones y acelera la convergencia en direcciones relevantes.

Actualización del momento

$$v_t = \gamma \cdot v_{t-1} + \eta \cdot \nabla L_{total}(\theta)$$

Actualización de pesos

$$\theta = \theta - v_t$$

aquí tenemos;

- γ es el factor de momentum
- η es la tasa de aprendizaje
- v_t es la velocidad acumulada del gradiente

3.6.3. Tasa de Aprendizaje inicial (ILR)

El Initial Learn Rate es un hiperparámetro clave en el entrenamiento de modelos de aprendizaje automático, ya que determina la magnitud de los ajustes que se realizan en los pesos del modelo durante cada iteración (o paso de entrenamiento). Estos ajustes se hacen en respuesta al error calculado, que es la pérdida (loss), y son guiados por el gradiente de la función de pérdida con respecto a los pesos.

Capítulo 4

Metodología

4.1. Software

MATLAB es un entorno de programación y un lenguaje de alto nivel. Es ampliamente utilizado en física, matemáticas, ingeniería y otras ciencias debido a sus capacidades para el cálculo numérico, el análisis de datos, la visualización y el desarrollo de algoritmos. MATLAB incluye una amplia gama de herramientas y funciones que facilitan el trabajo en diversas áreas, como el procesamiento de señales, la modelización matemática, la simulación y la ciencia de datos. [29].

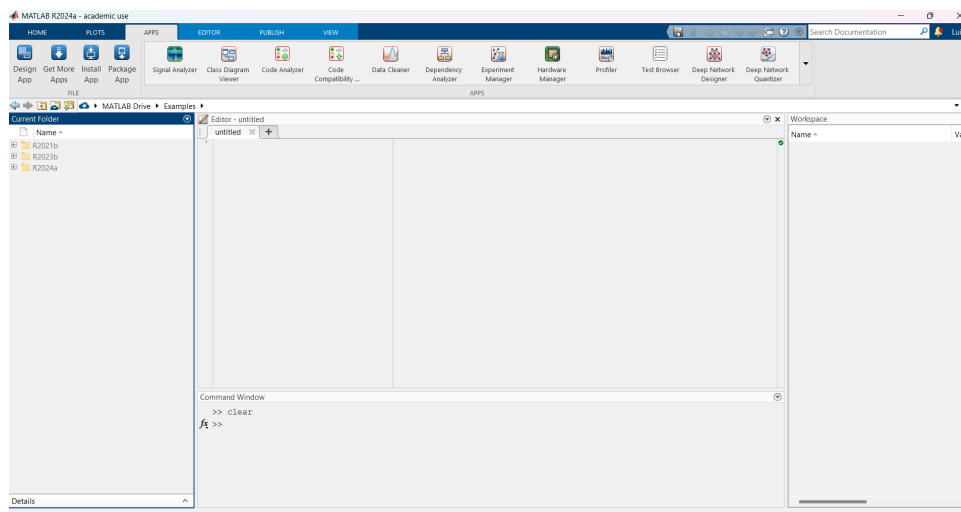
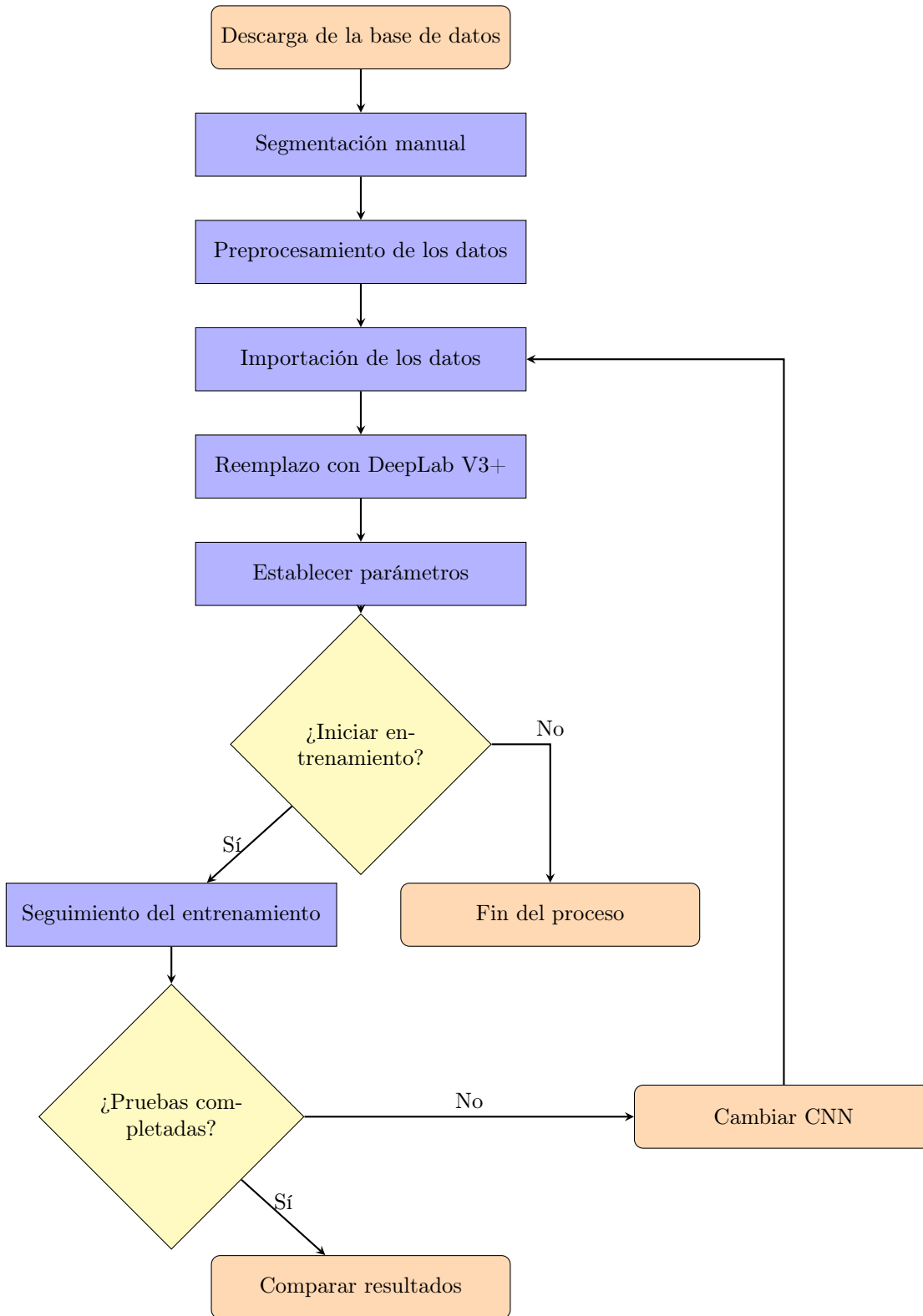


Figura 4.1: Interfaz principal de MATLAB

4.2. Diagrama de flujo

La metodología seguida durante este trabajo se resume en el siguiente diagrama. Posteriormente en las siguientes secciones se explica detalladamente cada paso.



4.3. Base de datos

La base de datos utilizada es The mini-MIAS una base que tiene alrededor de 300 imágenes en color gris de un tamaño de 1024×1024 píxeles. De las 300 imágenes, solo usamos un total de 89 de nuestro interés [30]. En la base se puede encontrar la información acerca del diagnostico de las mamografías que a continuación se muestran:

- Carácter del tejido de fondo.
 - Grasa.
 - Grasa glandular.
 - Densa glandular.
- Clase de anomalía presente.
 - Diferentes tipos de nodos.
 - Calcificaciones.
 - Asimetría.
 - Distorsión arquitectónica.
- Gravedad de la anomalía.
 - Benigna.
 - Maligna.
- Ubicación de la anomalía.
 - Coordenadas (x,y).
 - Radio aproximado de la anomalía.

Para el trabajo presente solo se usaron mamografías con calcificaciones y nodos.

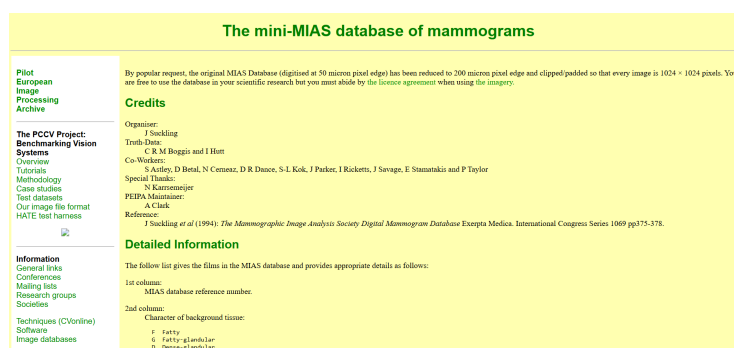


Figura 4.2: Base de datos mini-MIAS [30]

4.4. Preprocesamiento de las imágenes

Con la aplicación "Image Labeler" se etiquetaron de manera manual las imágenes de la base [31]. Se usaron 4 etiquetas diferentes para la segmentación; Fondo, Mama, Calcificaciones y Anomalías (Figura 4.3). En este último, solo se consideren las masas circunscritas, masas espiculadas y masas no definidas. De manera predeterminada en la aplicación se guardan las imágenes, etiquetas de píxeles (mascara de segmentación; Una mascara de segmentación es una matriz que

contiene la información de clase para cada pixel de una imagen) e información de las clases en un almacén de datos [32].

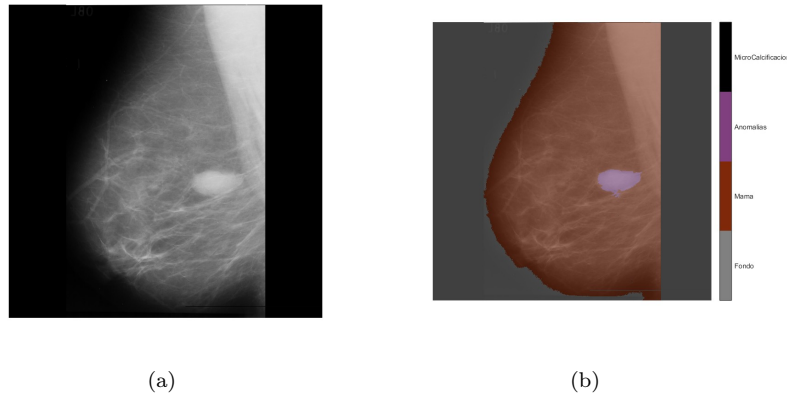


Figura 4.3: Imagen orinal. (b) Imagen original con su mascara de segmentación

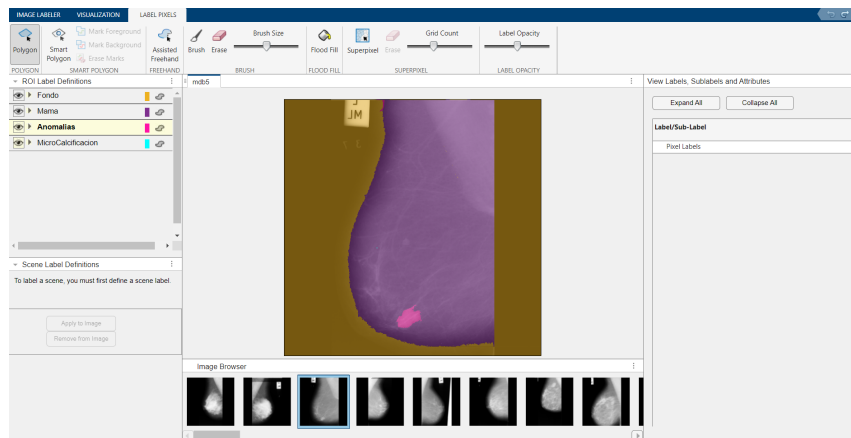


Figura 4.4: Interfaz de la aplicación "Image Labeler"

Después de que se guarda e importan las imágenes al espacio de trabajo, se separan las imágenes de su mascara de segmentación. En MATLAB se utilizan funciones para trabajar con almacenes de datos, como ImageDatastore [33] y pixelLabelDatastore [34], que están diseñadas específicamente para manejar colecciones de imágenes y etiquetas de píxeles (En este caso las mascaras de segmentación) respectivamente. En el apéndice A se muestra el código usado.

Redimensionamiento de las imágenes

El redimensionamiento de las imágenes implica perdida de calidad en la imagen, lo ideal es usar las imágenes en su calidad original de 1024×1024 , pero también queremos conocer el comportamiento de las redes con imágenes en baja resolución, entonces vamos a redimensionar las imágenes a 299×299 pixeles.

Para esto se tomaron en cuenta diferentes métodos para el redimensionamiento y se opto por el que obtuvo mejores resultados de manera estadística. Se consideran los 3 siguientes métodos [35];

- Interpolación bilineal:

La interpolación bilineal identifica los valores de los 4 píxeles vecinos más cercanos en la imagen original que rodeen la posición mapeada. La imagen resultante se obtiene mediante una combinación lineal de los valores de píxeles vecinos en la imagen original.

- Interpolación del vecino más cercano (nearest):

Este método asigna a cada nuevo píxel el valor del píxel más cercano de la imagen original, de manera similar esto sucede con las etiquetas de los píxeles. Es un método simple y rápido que no introduce nuevos valores, sino que reutiliza los valores existentes.

- Interpolación bicúbica

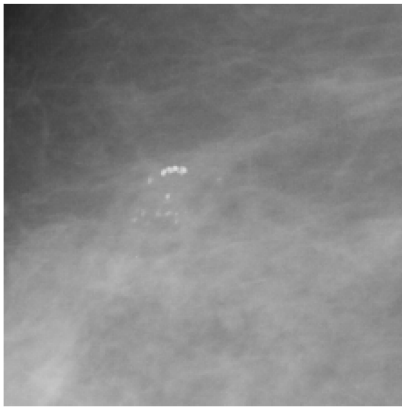
Se basa en el cálculo del valor de un píxel interpolado utilizando los valores de una vecindad de 16 píxeles (4x4) alrededor del píxel objetivo. Se ajustan curvas cúbicas para cada dimensión (horizontal y vertical) utilizando los valores de los píxeles vecinos.

En la figura 4.3 observamos los tres métodos de interpolación aplicados en una imagen, en este caso observamos que la zona de interés es una pequeña calcificación que se localiza en la parte central de la imagen original se pueden ver pequeños puntos blancos, con un contraste entre calcificación y mama significativo. En la figura c se observa que el contraste disminuye y la diferencia entre mama y calcificación es menos notoria esto porque el método suaviza los píxeles lo que implica pérdida de información importante. En la figura b y d se ve mejor la calcificación, pero hay un mejor contraste y detalle en la figura b.

Los resultados anteriores nos indican un mejor funcionamiento de la interpolación nearest para este problema. Con esto, se optó por usar el método para el proceso de redimensionamiento de las imágenes y etiquetas.

Un paso extra para las imágenes es pasar de un canal disponible (gris) a un canal triple (RGB). El proceso para obtener el canal RGB es solo triplicar el único canal disponible, como se espera la realidad es que no obtenemos un RGB, simplemente. Este paso es muy importante porque la mayoría de CNN existentes requieren imágenes con tres canales para su uso.

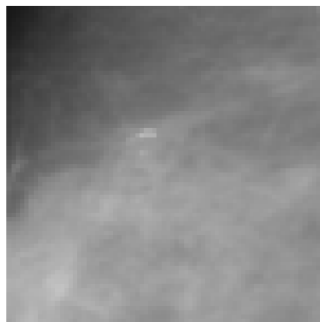
Posteriormente ambas bases se unen para crear una nueva base de datos que contiene las imágenes con canal tripe y sus etiquetas redimensionadas.



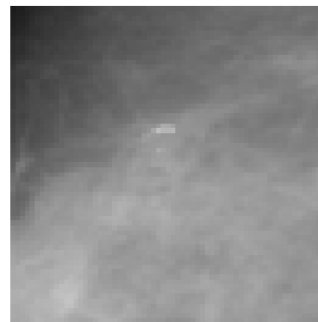
(a) Imagen original



(b) Nearest



(c) Bilineal



(d) Bicubica

Figura 4.5: a) Muestra parte de una imagen con 1024×1024 píxeles tomada de la base de datos. b) Imagen redimensionada a 299×299 píxeles con el método de interpolación nearest. c) Imagen redimensionada con el método de interpolación bilineal. d) Imagen redimensionada con el método de interpolación bicubica

4.4.1. Configuración de las nuevas capas

- Se crea una nueva instancia DeepLab V3+ especificando; el tamaño de entrada para las imágenes, el numero de clases para las etiquetas de píxeles y la red neuronal usada. [39]
- Se crea una nueva capa de entrada utilizando la función `imageInputLayer` [40] y se reemplaza en la red usada [41]. Parámetros usados:
 1. Tamaño de las imágenes de entrada = $1024 \times 1024 \times 3$; Las imágenes ya procesadas tienen un tamaño de 1024×1024 píxeles y 3 canales equivalentes al RGB.
 2. Normalización = `rescale-zero-one`; Cambia la escala de la entrada para que entrada ahora se encuentre en el intervalo $[0, 1]$ y así reducir carga computacional.
 3. Valor mínimo y máximo para reescalar = 0 a 255; Valores típicos para imágenes en escala de grises. Este valor se usa en la normalización

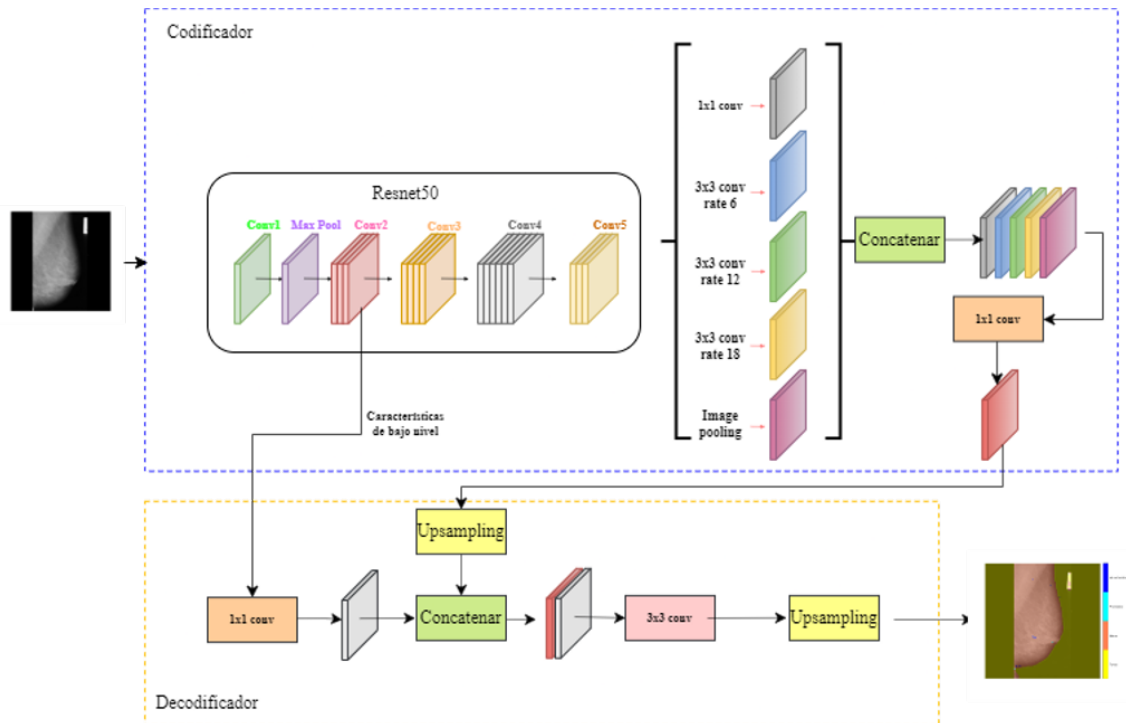


Figura 4.6: Arquitectura codificador-decodificador de Resnet + DeepLabV3+

Se configuran de la misma manera las nuevas capas para ambas CNNs.

4.5. Entrenamiento de las CNNs

Para trabajar de manera efectiva y controlada en cada una de las etapas de entrenamiento y resultados, se dividen los datos de la siguiente manera; 0.75 para el entrenamiento, 0.2 para prueba y 0.05 para la validación de los datos. Se utilizan las opciones de entrenamiento `trainingOptions` para definir varios parámetros que controlan el proceso de entrenamiento de la red neuronal [42], a continuación se muestran algunos de los parámetros más importantes para el entrenamiento de la red:

- `'sgdm'` (Stochastic Gradient Descent with Momentum): Algoritmo de optimización utilizado para actualizar los pesos de la red durante el entrenamiento.
- `InitialLearnRate` = Tasa de aprendizaje inicial para el algoritmo de optimización.
- `L2Regularization` = Parámetro de regularización L2, que ayuda a prevenir el sobreajuste penalizando grandes pesos en la red.
- `ValidationData` = Conjunto de datos utilizado para validar el rendimiento del modelo durante el entrenamiento.
- `ValidationFrequency` = Frecuencia con la que se realiza la validación (cada 30 iteraciones).
- `MaxEpochs` = Número máximo de épocas para entrenar la red.
- `MiniBatchSize` = Tamaño del mini-lote, es decir, el número de ejemplos de entrenamiento procesados antes de actualizar los pesos de la red.

Para conocer los valores usados para cada parámetro ver el Apéndice 1.

Después de definir las opciones de entrenamiento, se comienza el entrenamiento de la red. Durante el entrenamiento se gráfica la pérdida de entrenamiento y pérdida de la validación. Al finalizar el entrenamiento se guarda la red y se procede a repetir el proceso para entrenar la red con nuevos parámetros.

Importante mencionar que se realizaron diferentes configuraciones de entrenamiento. Esto con el propósito de obtener los mejores resultados con diferentes configuraciones.

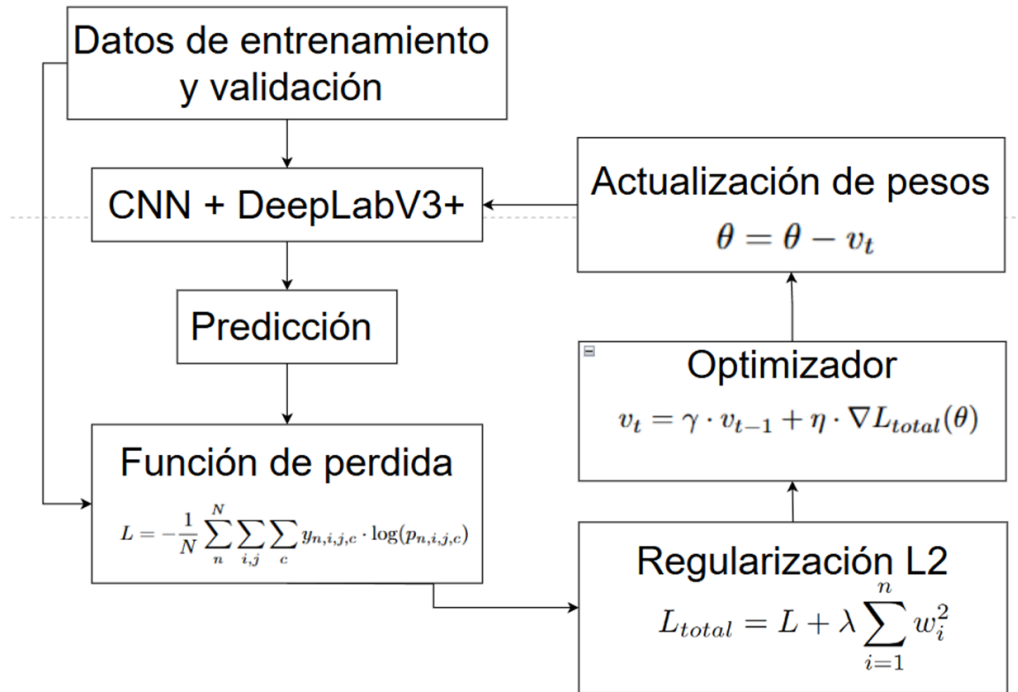


Figura 4.7: Diagrama de entrenamiento de las CNNs

Capítulo 5

Resultados

5.1. Eficiencia de las CNNs

En esta sección se presentan los resultados del análisis de segmentación de mamografías utilizando el modelo ResNet50, evaluando el impacto de la regularización L2 y la tasa de aprendizaje inicial (ILR) en el desempeño de la segmentación para imágenes de 1024×1024 y 299×299 píxeles respectivamente. Para ello, se analizan métricas; precisión (Accuracy) y la intersección sobre la unión (IoU) en las dos partes de interés de la mama: anomalías y calcificaciones.

5.1.1. Configuración de las pruebas

Se realizaron un total de 6 configuraciones con los parámetros variables, 10 pruebas por configuración. Nos enfocamos en valores típicos de los dos parámetros; regularización L2 y Tasa de aprendizaje inicial.

	Regularización L2		
ILR	0.001	0.005	0.01
0,01	✓	✓	✓
0,1	✓	✓	✓

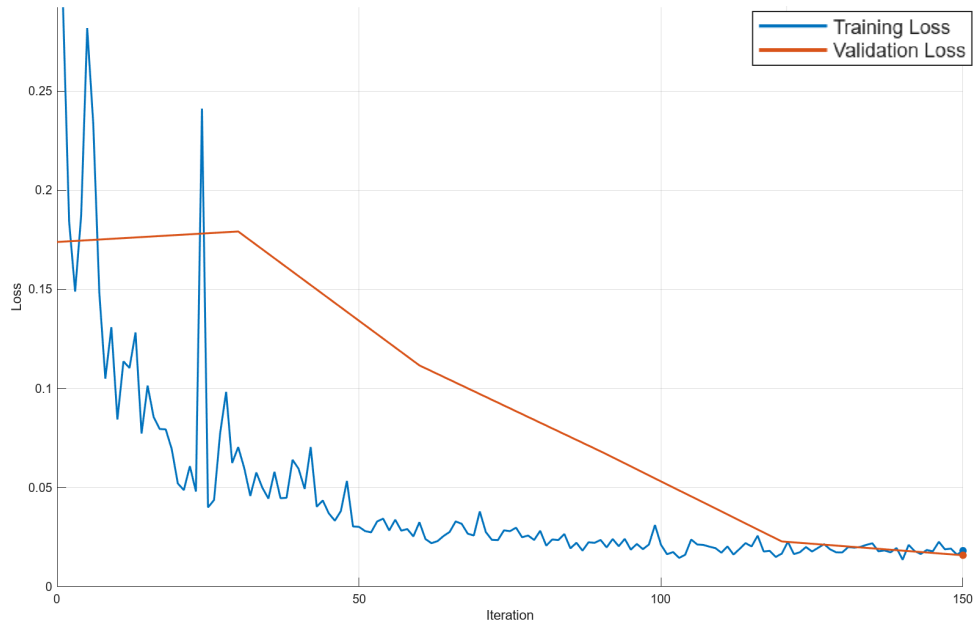
Tabla 5.1: Configuraciones de los parámetros variables durante las pruebas de entrenamiento.

5.2. Curvas de Pérdida del entrenamiento

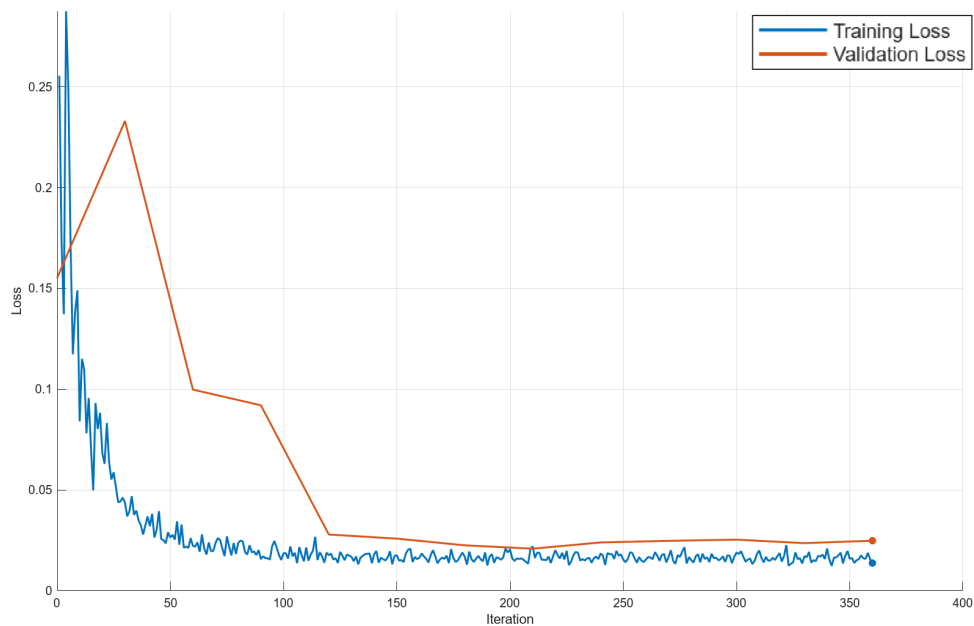
Observe la Figura 5.2 y la Fig 5.1. Ambas figuras corresponde a la evolución en la curva de pérdida del entrenamiento y validación para Resnet 50 con 299×299 y 1024×1024 píxeles respectivamente.

La evolución de la pérdida durante el entrenamiento de la Resnet 50 con las imágenes de reducidas (5.1) disminuye progresivamente con algunas oscilaciones mas grandes en las primeras iteraciones. Por otro lado, la pérdida de validación comienza con una ligera tendencia ascendente, pero posteriormente disminuye de manera más pronunciada. Esto indica que de inicio el modelo tiene dificultades para generalizar, pero con más iteraciones logra reducir el error en los datos de validación. La convergencia entre la perdida del entrenamiento y la validación se nota mas en la Figure 5.2b que corresponde al valor de tasa de aprendizaje inicial igual a0,01, esta convergencia indica un mayor control en el sobreajuste y subajuste.

Las curvas de la Figura 5.2 que corresponde al entrenamiento con imágenes de 1024×1024 píxeles tienen un comportamiento similar a las del entrenamiento con imágenes reducidas, tanto en entrenamiento como en validación. Con una diferencia en la curva de validación que sigue de cerca a la pérdida de entrenamiento desde el inicio, lo que puede indicar una mejor estabilidad en el proceso de aprendizaje. De igual manera que en el entrenamiento con imágenes reducidas, presenta una divergencia pequeña al final de la curva con la tasa de aprendizaje inicial igual 0.1.

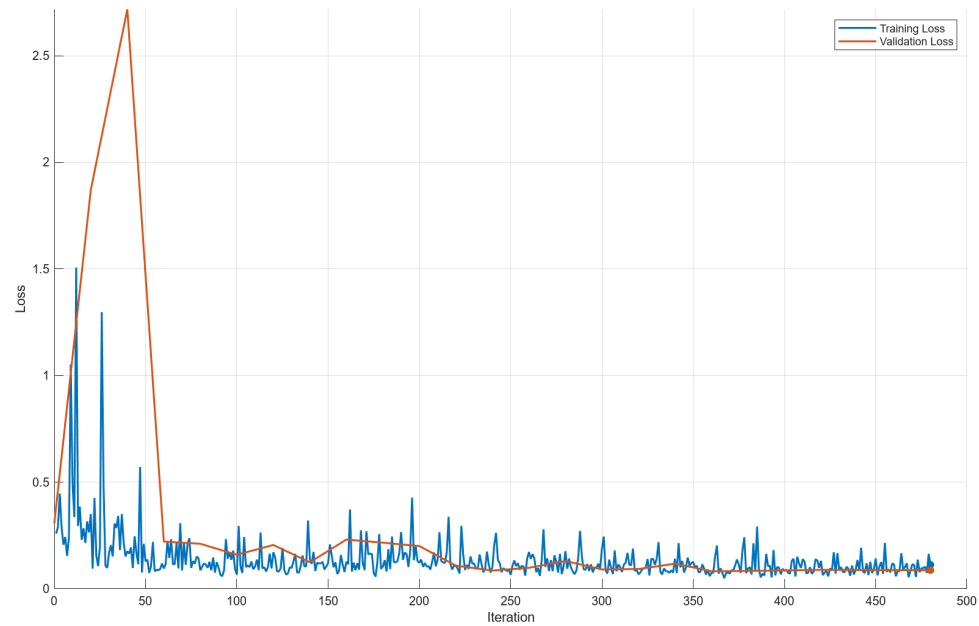


(a)

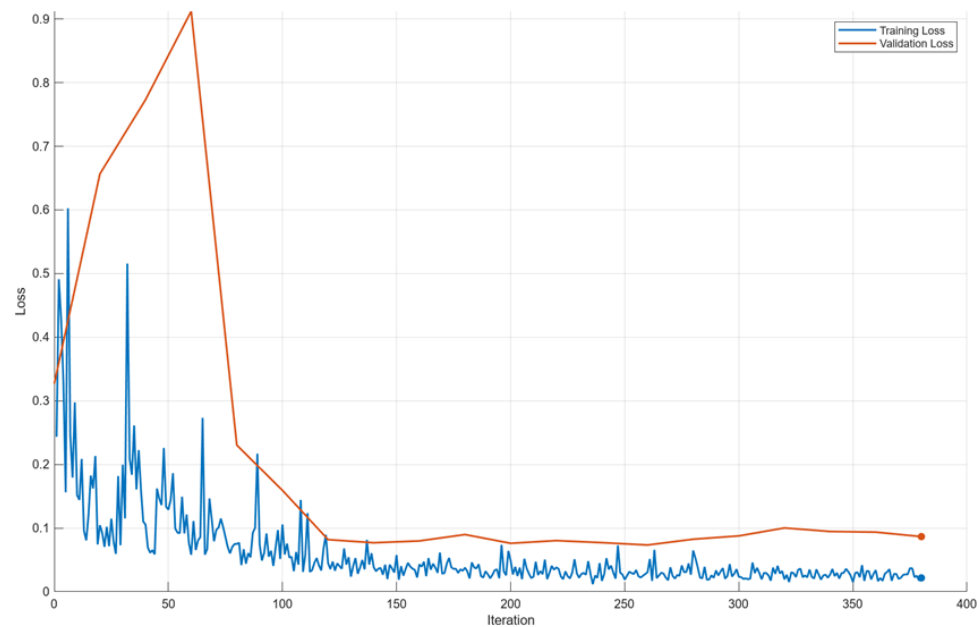


(b)

Figura 5.1: Curvas de pérdida en el entrenamiento de Resnet 50 con imágenes de 299×299 píxeles y tasa de aprendizaje inicial de a) 0,01 y b) 0,1



(a)



(b)

Figura 5.2: Curvas de pérdida en el entrenamiento de Resnet 50 con imágenes de 1024×1024 píxeles y tasa de aprendizaje inicial de a) 0,01. b) 0,1

Hemos mencionado el sobreajuste y subajuste, pero ¿qué es?

- El sobreajuste ocurre cuando el modelo se ajusta demasiado a los datos de entrenamiento, capturando ruido o detalles específicos que no generalizan bien a nuevos datos.
- El subajuste ocurre cuando el modelo no se ajusta a los datos de entrenamiento, capturando pocas características, perdiendo información en nuevos datos.

5.3. Métricas

Una métrica es un valor numérico que resume el desempeño del modelo entrenado en relación con el conjunto de datos de prueba y validación. Estas métricas se calculan comparando las predicciones del modelo con los valores reales. Se utilizan a continuación dos métricas;

5.3.1. Precisión (Accuracy)

Accuracy es una métrica utilizada para evaluar la segmentación que solo considera verdaderos positivos (VP), falsos negativos (FN) y el total de píxeles (TP) para cada clase. La métrica tiene resultados entre 0 y 1, entre mas cerca de 1 mejor es la precisión.

$$Accuracy = \frac{VP}{TP + FN} \quad (5.1)$$

En la figura 5.3 y figura 5.4 se muestran los resultados obtenidos para las calcificaciones y nodos respectivamente con la métrica accuracy. Veamos primero 5.3, en esta figura se observa el comportamiento de Resnet-50 con las 4 configuraciones principales y su comportamiento con diferentes valores de Regularización L2. En los resultados no hay una clara diferencia entre las configuraciones con imágenes de 1024×1024 y las de 299×299 donde la configuración de Resnet 50 con imágenes de 299×299 y un tasa de aprendizaje inicial de 0.01 es mejor en la segmentación de calcificaciones con todo los valores de regularización L2.

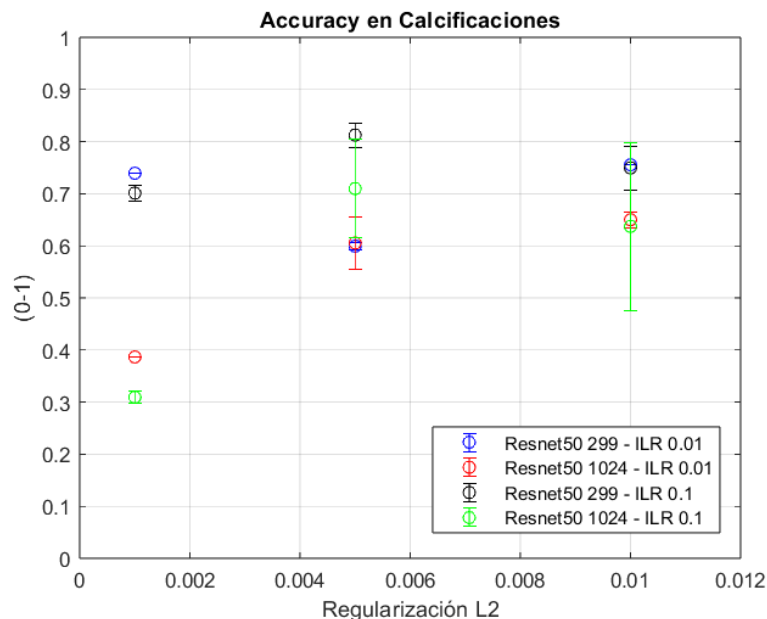


Figura 5.3: Precisión en las calcificaciones para las diferentes configuraciones.

Ahora, para el caso de los nodos en la figura 5.4 podemos notar como las configuraciones con imágenes de 1024×1024 tienen mejores resultados, a excepción de algunas configuraciones, donde no hay una clara diferencia con los resultados del entrenamiento con imágenes de 299×299 y tasa de aprendizaje inicial de 0.1.

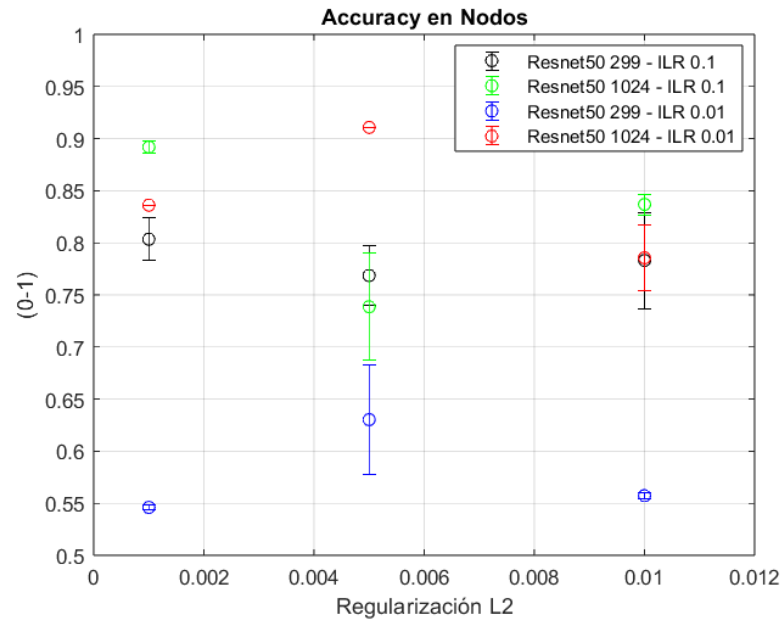


Figura 5.4: Precisión en los nodos para las diferentes configuraciones

5.3.2. Intersección sobre la unión (IoU)

IoU es una métrica que mide la superposición entre dos regiones: la región predicha por un modelo y la región real. A diferencia de Accuracy en esta métrica se penalizan también los falsos positivos(FN). Al igual que Accuracy los resultados caen entre 0 y 1, con mejores resultados si el IoU es mas cercano a 1.

$$IoU = \frac{VP}{TP + FN + FP} \quad (5.2)$$

Veamos que los resultados obtenidos en esta métrica son muy inferiores a los que podríamos esperar y al contrario de los resultados obtenidos en la precisión el modelo entrenado con imágenes de 299×299 pixeles y una tasa de aprendizaje inicial igual a 0.01 obtuvo mejores resultados en la mayoría de configuraciones para ambos datos de interés (Calcificaciones y nodos).

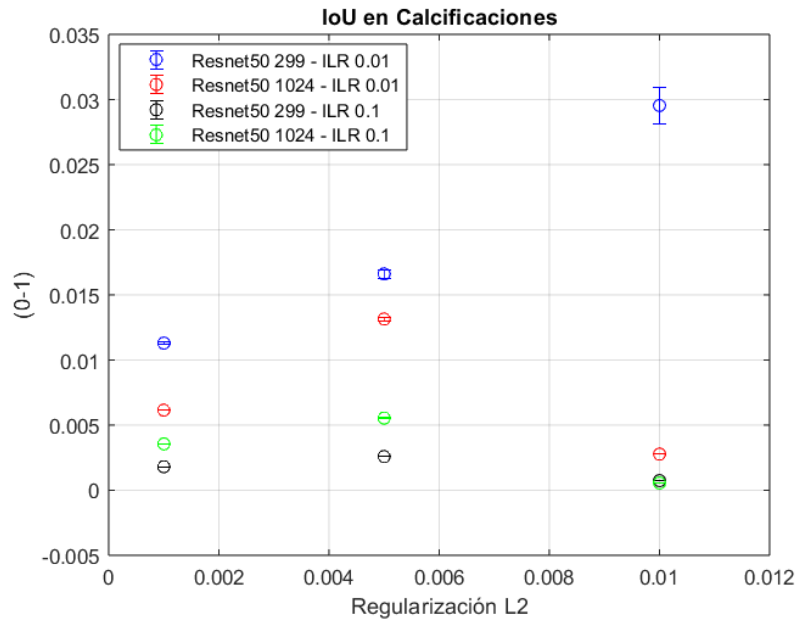


Figura 5.5: intersección sobre la unión en las calcificaciones para las diferentes configuraciones.

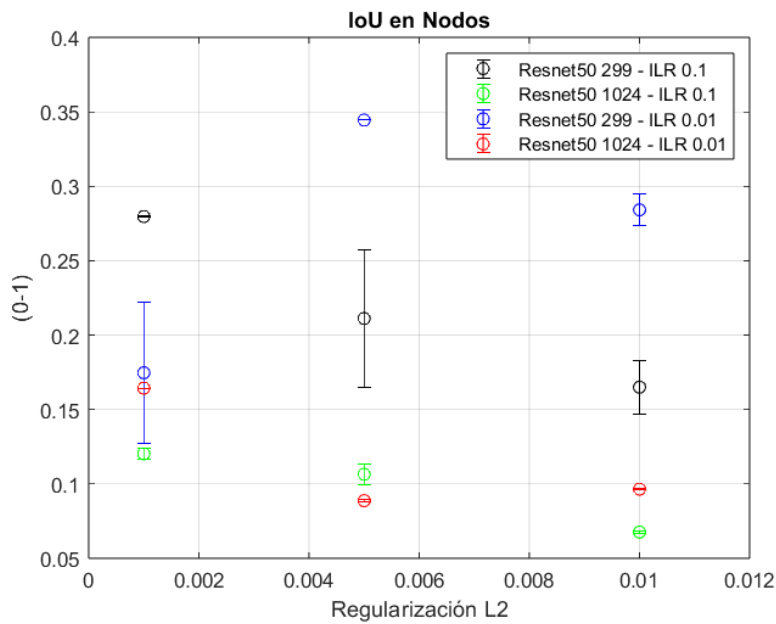


Figura 5.6: intersección sobre la unión en los nodos para las diferentes configuraciones.

¿Que son los VP, FN y FP?

Cada imagen segmentada con el modelo entrenado crea una mascara que da valores a cada pixel, estos pixeles se pueden nombrar como verdaderos positivos, falsos negativos u otro dependiendo de su comparación.

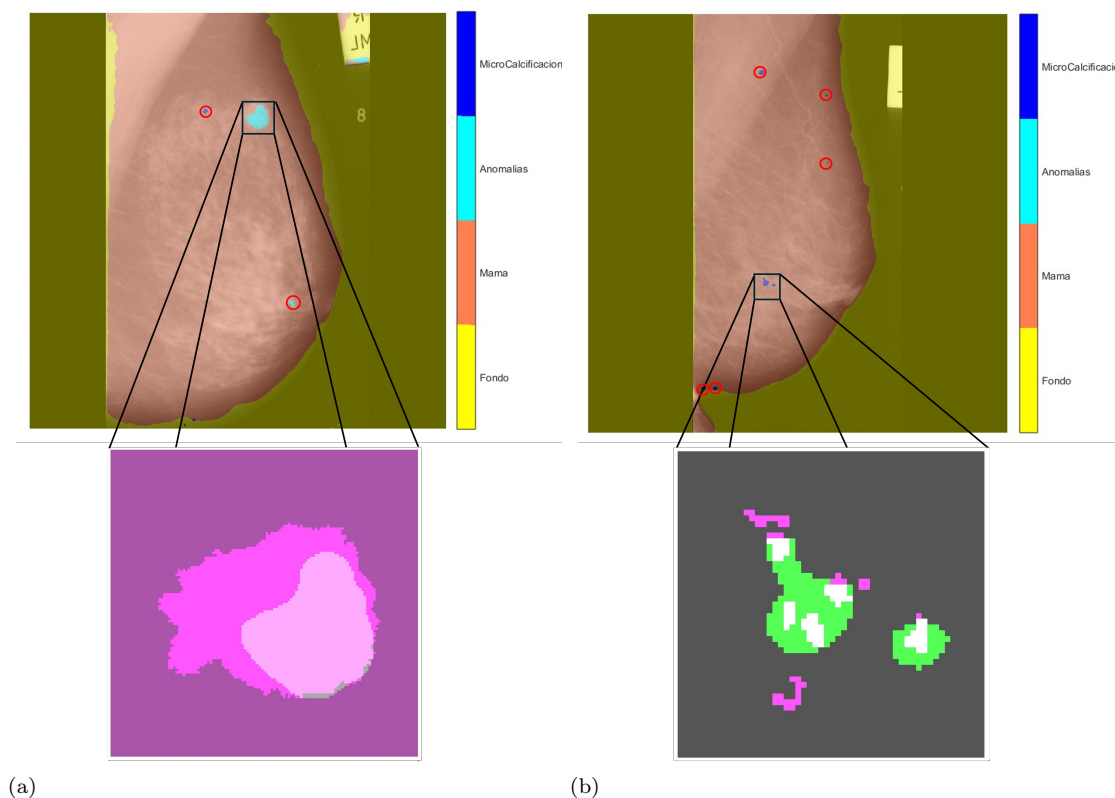


Figura 5.7: Imagen segmentada donde la mama presenta a) una anomalía mal definida b) calcificaciones:

Para entender mejor, consideremos un ejemplo en el que se necesita obtener las métricas para las calcificaciones.

- Verdaderos Positivos(VP): Son los pixeles segmentados como calcificaciones y que en realidad si corresponden a una calcificación. De manera mas sencilla, son los pixeles segmentados correctamente.
- Falsos Negativos(FN): Son los pixeles segmentados como calcificaciones, pero en realidad corresponden a otra clase.
- Falsos Positivos(FP): Son los pixeles segmentados como una clase diferente a la calcificación, pero en realidad son calcificaciones.

La Figura 5.7 muestra dos ejemplos en los que se pueden observar dos casos que tienen Verdaderos positivos, Falsos negativos y Falsos positivos. La figura 5.7a presenta una anomia de tipo nódulo, de tipo mal definida. Las partes encerradas en un círculo rojo son falsos positivos de la anomalía. Si bien la imagen 5.7a completa es solo la mascara de segmentación obtenida, se muestra también una zona aumentada, que es parte de la superposición por clase verdadera. En esta parte se superpone la capa de segmentación verdadera y la capa de segmentación hecha por el modelos. Estos datos son usados en el IoU 5.7a, donde, el tono rosa claro corresponde a verdaderos positivos, el rosa mas intenso a los falsos negativos y una pequeña zona gris con falsos positivos. De la misma manera, en la figura 5.7b, se muestran los falsos positivos encerrados en pequeños círculos rojos y en el parte aumentado el tono blanco corresponde a verdaderos positivos, el verde a falsos positivos y en rosa los falsos negativos.

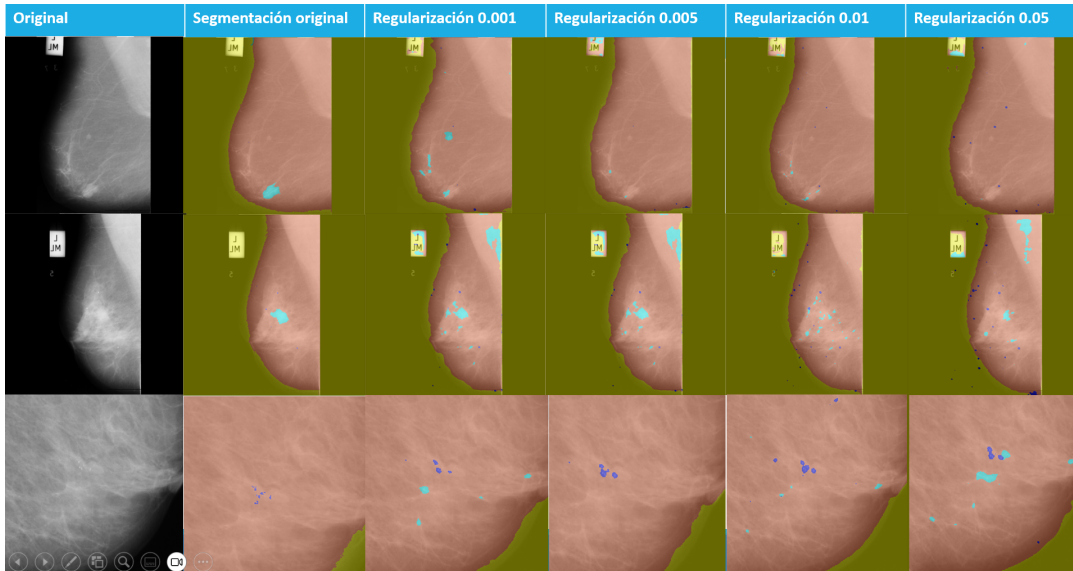


Figura 5.8: Segmentación generado por el modelo entrenado con imágenes reducidas

5.3.3. Resultados globales

Los resultados globales consideran la segmentación de las cuatro etiquetas con las que se entrenaron ambas redes neuronales convolucionales (CNNs): Mama, calcificaciones, Anomalías y Fondo. En general, se observan resultados con una alta precisión (Precisión $>80\%$). Sin embargo, la precisión global puede no reflejar de manera adecuada el rendimiento en las categorías menos frecuentes, como son las calcificaciones y las anomalías, debido a que los resultados se ven influenciados por el tamaño relativo de cada clase. Por otro lado, la precisión media asigna un valor igual para todas las etiquetas, proporcionando una evaluación más equilibrada del desempeño de la segmentación en cada categoría, independientemente de su tamaño. Veamos las tablas 5.2 y 5.3 que corresponden a la precisión global y media de las diferentes configuraciones de Resnet50 con imágenes de 1024×1024 y Resnet 50 imágenes de 299×299 píxeles, respectivamente. Recordemos que para cada configuración se realizaron 8 pruebas, es por esto que podemos asociarles un error a los datos.

	Taza de aprendizaje inicial			
	0.01	0.1	0.01	0.1
Regularización L2	Precisión global		Precisión media	
0.001	0.9163 ± 0.000043	0.9174 ± 0.0011	0.82165 ± 0.0181	0.81009 ± 0.0979
0.005	0.9166 ± 0.0011	0.8750 ± 0.000117	0.839335 ± 0.019	0.79185 ± 0.00428
0.01	0.8901 ± 0.000524	0.9137 ± 0.0027	0.81617 ± 0.0337	0.82755 ± 0.0026

Tabla 5.2: Tabla de precisión media y global de las pruebas con imágenes de 1024×1024 píxeles.

Regularización L2	Taza de aprendizaje inicial			
	0.01	0.1	0.01	0.1
	Precisión global		Precisión media	
0.001	0.946125 ± 0.0138	0.872675 ± 0.0759	0.795985 ± 0.0034	0.80567 ± 0.0409
0.005	0.949085 ± 0.0149	0.908005 ± 0.0559	0.813205 ± 0.0145	0.74563 ± 0.1073
0.01	0.94946 ± 0.0146	0.884075 ± 0.0808	0.797645 ± 0.0070	0.82561 ± 0.0535

Tabla 5.3: Tabla de precisión media y global de las pruebas con imágenes de 299×299 píxeles

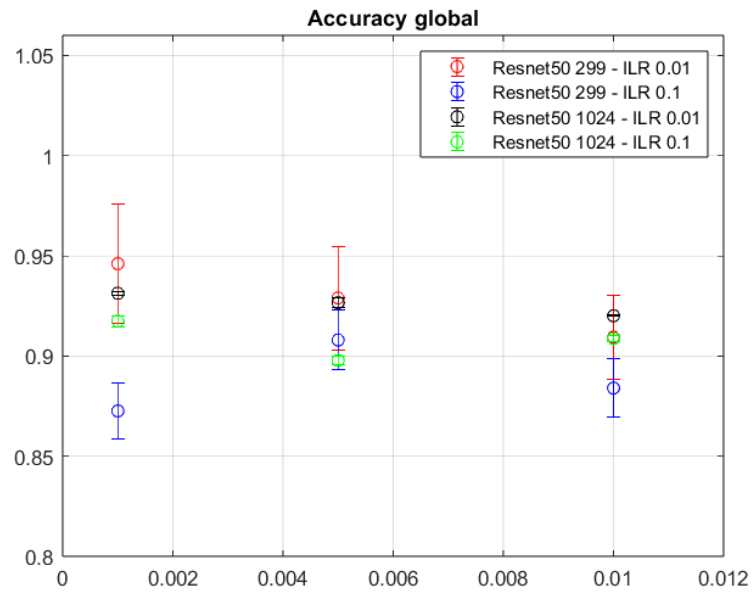


Figura 5.9: Gráfica del accuracy global para ResNet con imágenes de 299×299 y 1024×1024 píxeles

Capítulo 6

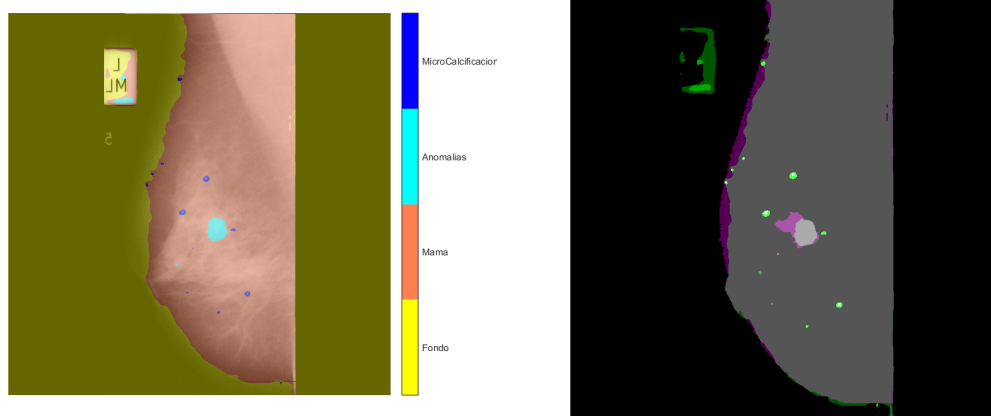
Discusión y conclusiones

Uno de los objetivos era determinar cómo el tamaño de las imágenes puede afectar el rendimiento de segmentación. Además, comparar cómo afectan las diferentes configuraciones entre la tasa de aprendizaje inicial y la regularización L2, con el fin de evaluar la estabilidad y capacidad de aprendizaje con ambos tamaños de imágenes.

Los resultados obtenidos permitieron identificar tendencias clave sobre la influencia de la calidad de las imágenes en la segmentación de estructuras relevantes en mamografías proporcionando información útil para la optimización de los modelos de segmentación en tareas de detección médica.

Vemos pequeñas diferencias en las configuraciones con diferentes valores de tasa de aprendizaje inicial y de regularización L2, estas diferencias pueden significar grandes mejoras de entrenamiento, con más imágenes o tal vez con configuraciones de otros parámetros que nosotros mantuvimos constantes.

Observamos que reducir las imágenes puede mejorar la métrica del IoU pero que no necesariamente indica que los resultados sean mejores, simplemente es un efecto de tener menos cantidad de píxeles que pueden afectar el valor de esta métrica.



(a)

(b)

Figura 6.1: **a)** Muestra la segmentación de Resnet 50 con la siguiente configuración; Regularización L2 = 0.01 y ILR=0.01. **b)** Muestra una máscara de verdad de a)

En general observamos que reducir el tamaño de las imágenes no afecta drásticamente el desempeño del modelo, en los resultados de accuracy se observa una mínima diferencia entre el modelo entrenado con imágenes de 1024×1024 y el modelo entrenado con imágenes de 299×299 pixeles y no solo eso, también obtuvimos configuraciones donde la reducción de la imagen nos ayudo a evitar el sobreajuste de manera mínima a comparación de las imágenes originales. Este resultado se contrasta con los datos del artículo [26] donde se muestra que el cambio de la resolución en imágenes no afecta demasiado los resultados. Esto resultados son esenciales para comprender como las redes neuronales aprenden y mejorar estos algoritmos.

Conjunto de Datos	$\times \frac{1}{8}$	$\times 1$
Base de datos 1	0.587	0.618
Base de datos 2	0.718	0.794
Base de datos 3	0.729	0.787
Resnet50-Minimias	0.82561 ± 0.0535	0.839335 ± 0.019

Tabla 6.1: Resultados de la arquitectura Connected-UNets en distintos conjuntos de datos [26] y de la arquitectura Resnet50 con minimias

Por ultimo, podemos concluir que existen configuraciones que mejoran la segmentación de alguna parte en especifica pero tal vez lo ideal sea trabajar con datos por separado y así obtener los mejores resultados posibles, ademas, si es posible siempre trabajar con imagen de una buena calidad, esto para evitar la perdida de información y no solo eso también usar bases de datos mas grandes, al menos mas grande que la base que se uso en este trabajo.

Apéndice A

A.1. Preprocesamiento y entrenamiento

El código se divide en dos scripts, El primero realiza todo el preprocesamiento de los datos y su almacenamiento. El segundo realiza la modificación de la capa de entrada y el entrenamiento de la CNN. Esta separación del código es esencial para tener las mismas condiciones en cada configuración. Para el código se usan algunas funciones no incluidas en el software. Algunas funciones son reutilizadas de trabajos previos en MATLAB y otras son funciones personalizadas.

Preprocesamiento de los datos

```
1     %% Preprocesamiento de los datos.
2     % Carga las imagenes y sus etiquetas.
3     load('gTruth.mat')
4     % Se establece el tamaño de las imagenes para su preprocesamiento.
5     TamImagen = [1024 1024 3];
6     % Se crea un datastore con las imagenes.
7     Dstore=imageDatastore(gTruth.DataSource.Source);
8     % Se crea un pixelLabelDatastore con la informacion de las etiquetas.
9     pixelDataStore = pixelLabelDatastore(gTruth);
10    % Define el mapa de colores que se encuentra en la funcion ColorMap.
11    MapaColores = ColorMap;
12    % Define el nombre de las clases para la segmentacion.
13    clases=gTruth.LabelDefinitions.Name;
14    % Realiza un conteo del numero de pixeles para cada clase.
15    tbl = countEachLabel(pixelDataStore);
16    % Determina la frecuencia de pixeles clase vs total de pixeles.
17    frecuencia = tbl.PixelCount/sum(tbl.PixelCount);
18    % procesa las imagenes con la funcion transform.
19    transfo_imagedatastore = transform(Dstore,@(img) preprocessImage(img,TamImagen)
20    );
21    % Procesa las etiquetas con la funcion auxiliar preprocesslabels.
22    transfo_pixeldatastore = transform(pixelDataStore, @(img)
23    preprocessLabels(img,TamImagen));
24    % Combina el imagedatastores y el pixelLabelDatastore.
25    datastore_Combinado = combine(transfo_imagedatastore,transfo_pixeldatastore);
26    % Divide los datos para el entrenamiento, validacion y prueba de la red.
27    numImages = numel(Dstore.Files);
28    numEntrenamiento = round(0.75*numImages);
29    numValidacion = round(0.05*numImages);
30    numPrueba = round(0.2*numImages);
31    % Guarda las imagenes en cada una de la variables correspondientes.
32    sIndices = randperm(numImages);
33    dsEntrenamiento = subset(dataStore_Combinado,sIndices(1:numEntrenamiento));
34    dsValidacion = subset(dataStore_Combinado,sIndices(numEntrenamiento+1:
35    numEntrenamiento+numValidacion));
36    dsPrueba = subset(dataStore_Combinado,sIndices(numEntrenamiento+numValidacion
37    +1:end));
```

```

35 % Aumenta los datos de entrenamiento con la funcion
    augmentDataForLVSegmentation.
36 dsEntrenamiento = transform(dsEntrenamiento,@(data)
    augmentDataForLVSegmentation(data));
37 dsValidacion=transform(dsValidacion,@(data1) augmentDataForLVSegmentation(data1
    ));
38 dsPrueba=transform(dsPrueba, @(data2) augmentDataForLVSegmentation(data2));

```

Entrenamiento de la red

```

1 %% Modificacion de la CNN con DeepLabV3+
2 % Cargamos las imagenes y sus datos de segmentacion
3 load('gTruth.mat')
4 % Tamano de entrada de las imagenes
5 TamImagen = [1024 1024 3];
6 % Definimos el nombre de las clases para la segmentacion
7 clases=gTruth.LabelDefinitions.Name;
8 % Cargamos los datos procesados
9 load("dataStore_Combinado.mat")
10 load("dsEntrenamiento.mat")
11 load("dsPrueba.mat")
12 load("dsValidacion.mat")
13 load("tbl.mat")
14 % Modificacion de la capa de entrada de la red
15 %Numero de clases
16 numClases = 4;
17 %Creamos una capa DeepLab V3+ con la red "resnet50" y los datos de las
18 % imagenes
19 networkDeep = deeplabv3plus(TamImagen,numClases, "resnet50");
20 %Modificamos la entrada de la red.
21 inputlayer = imageInputLayer(TamImagen, ...
22     Normalization="rescale-zero-one", ...
23     Min=0, ...
24     Max=255, ...
25     Name="inputlayer");
26 % Reemplazamos la capa original con la nueva capa de entrada
27 network = replaceLayer(networkDeep,networkDeep.Layers(1, 1).Name,inputlayer);
28
29 % Datos sobre numero de pixeles por etiqueta
30 imageFreq = tbl.PixelCount ./ tbl.ImagePixelCount;
31 classWeights = median(imageFreq) ./ imageFreq;
32
33 %% Opciones de entrenamiento
34 Opciones = trainingOptions("sgdm",...
35     LearnRateSchedule="piecewise",...
36     LearnRateDropFactor=0.1,...
37     Momentum=0.9,...
38     InitialLearnRate=1e-1,... % Variable
39     L2Regularization=0.005,... % Variable
40     ValidationData=dsValidacion,...
41     ValidationFrequency=20, ...
42     Plots="training-progress",...
43     MaxEpochs=50,...
44     MiniBatchSize=4,...
45     Shuffle="every-epoch",...
46     CheckpointPath=tempdir,...
47     VerboseFrequency=10,...
48     ValidationPatience=6);
49 IniciarEntrenamiento = true;
50 if IniciarEntrenamiento
51     [net,informacio_Proceso] = trainnet(dsEntrenamiento,network,@(Y,T) modelLoss(Y,
52         T,classWeights),Opciones);
53 end

```

A.2. Funciones

Establece el mapa de colores

```

1     function cmap = ColorMap()
2     % Definimos el mapa de colores.
3
4     cmap = [
5         128 128 128 % Fondo
6         128 40 10  % Mama
7         128 64 128  % Anomalias
8         1 0 0];    % Calcificacion
9
10    % Normalizacion entre [0 1].
11    cmap = cmap ./ 255;
12    end

```

Redimensiona las imágenes y triplica el canal gris

```

1     %% preprocessImage(imagenes, TamImagenNew)
2     function out = preprocessImage(imagenes, TamImagenNew)
3
4         TamImagenNew = TamImagenNew(1:2);
5         imagenes = imresize(imagenes, TamImagenNew);
6
7         if size(imagenes,3) == 1
8             imagenes = repmat(imagenes, [1 1 3]);
9         end
10
11        out = {imagenes};
12
13    end
14    % Copyright 2023 The MathWorks, Inc.

```

Redimensiona las etiquetas

```

1     function out = preprocessLabels(labels, targetSize)
2     % Copyright 2023 The MathWorks, Inc.
3
4         targetSize = targetSize(1:2);
5         labels = imresize(labels{1}, targetSize, "nearest");
6
7         out = {labels};
8
9     end

```

Aumento de los datos

```

1     function out = augmentDataForLVSegmentation(data)
2     % Copyright 2023 The MathWorks, Inc.
3
4         img = data{1};
5         labels = data{2};
6         inputSize = size(img, [1 2]);
7
8         tform = randomAffine2d(...
9             Rotation=[-5 5],...
10            XTranslation=[-10 10],...
11            YTranslation=[-10 10]);
12

```

```
13 sameAsInput = affineOutputView(inputSize,tform,BoundsStyle="sameAsInput");
14 img = imwarp(img,tform,"linear",OutputView=sameAsInput);
15 labels = imwarp(labels,tform,"nearest",OutputView=sameAsInput);
16
17 out = {img,labels};
18
19 end
```

Calculo de la perdida loss durante entrenamiento

```
1 function loss = modelLoss(Y, T, classWeights)
2 % Convertir los pesos de las clases a un arreglo profundo (dlarray)
3 weights = dlarray(classWeights, "C");
4
5 % Crear una mascara para los valores no faltantes
6 mask = ~isnan(T);
7
8 % Reemplazar NaNs en las etiquetas con cero
9 T(isnan(T)) = 0;
10
11 % Calcular la perdida de entropia cruzada ponderada con la mascara
12 loss = crossentropy(Y,T,weights,Mask=mask,NormalizationFactor="mask-included");
13 end
```

Bibliografía

- [1] Instituto Nacional de Estadística y Geografía. (2024). A propósito del Día Mundial de la Lucha contra el Cáncer de Mama, 2024. *INEGI*. Recuperado de https://www.inegi.org.mx/contenidos/saladeprensa/aproposito/2024/EAP_LuchaCMama24.pdf
- [2] Mejía Cortés, M. A. (2023). *Detección de cáncer en imágenes de mamografía usando aprendizaje profundo* (Trabajo de grado). Universidad de Ibagué.
- [3] Mera Giraldo, A. V. (2023). *Segmentación y visualización de bordes en cáncer de mama utilizando herramientas de inteligencia artificial* (Trabajo de grado). Universidad Autónoma de Occidente.
- [4] González-Bueno Puyal, J., Ríos Sánchez, B., & Sánchez Ávila, C. (2015). Detección y clasificación de tejidos anómalos en mamografías digitales mediante redes neuronales convolucionales. En *Libro de actas del Congreso INVE_MEM 2015*. Universidad Politécnica de Madrid.
- [5] Waschke, J., Koch, M., Kürten, S., Schulze-Tanzil, G., & Spittau, B. (2018). Tronco. En J. Waschke, M. Koch, S. Kürten, G. Schulze-Tanzil, & B. Spittau (Eds.), *Sobotta. Texto de anatomía* (pp. 27–56). Elsevier. <https://doi.org/10.1016/B978-84-9113-226-4.00002-3>
- [6] Drake, R. L., Vogl, A. W., Mitchell, A. W. M., Tibbitts, R. M., Richardson, P. E., & Horn, A. (2021). Tórax. En R. L. Drake et al. (Eds.), *Gray. Atlas de anatomía* (3.^a ed., pp. 61–131). Elsevier. <https://doi.org/10.1016/B978-84-9113-960-7.00003-6>
- [7] Martínez Jiménez, M., Alvarado Mendoza, M. del C., & Torres Muñoz, A. (2018). *Sistema de informes y registro de datos de estudios por imagen de mama (BI-RADS)* (5.^a ed.). *Anales de Radiología México*, 2(2), 105–112. Recuperado de <https://www.medigraphic.com/pdfs/anaradmex/arm-2018/arm182c.pdf>
- [8] Drake, R. L. (2024). *Gray. Anatomía para estudiantes* (5.^a ed.). Elsevier España, S.L.U.
- [9] Cruz Hernández, J. J., Rodríguez Sánchez, C. A., Del Barco Morillo, E., & Fonseca Sánchez, E. (2018). *Oncología clínica* (6.^a ed.). Elsevier España, S.L.U.
- [10] Zornoza, G., Arnáiz, J., & Lafuente, M. (2010). Clasificación radiológica y manejo de las lesiones mamarias mediante el sistema BI-RADS. *Clínica e Investigación en Ginecología y Obstetricia*, 37(6), 253–261. <https://doi.org/10.1016/j.gine.2010.08.002>
- [11] Bushong, S. C. (2010). Mamografía. En *Manual de radiología para técnicos* (9.^a ed., pp. 318–329). Elsevier.
- [12] Bushberg, J. T., Seibert, J. A., Leidholdt, E. M., & Boone, J. M. (2012). *The essential physics of medical imaging* (3rd ed.). Lippincott Williams & Wilkins.
- [13] Khan, F. M. (2010). *The physics of radiation therapy* (3rd ed.). Lippincott Williams & Wilkins.

- [14] Kittel, C. (2005). *Introduction to solid state physics* (8th ed.). Wiley.
- [15] Knoll, G. F. (2010). *Radiation detection and measurement* (4th ed.). Wiley.
- [16] Krane, K. S. (2012). *Modern physics* (3rd ed., pp. 92–93). Wiley.
- [17] Krane, K. S. (2012). *Modern physics* (3rd ed., pp. 78–79). Wiley.
- [18] Compton, A. H. (1923). A quantum theory of the scattering of X-rays by light elements. *Physical Review*, 21(5), 483–502. <https://doi.org/10.1103/PhysRev.21.483>
- [19] Eisberg, R., & Resnick, R. (1985). *Quantum physics of atoms, molecules, solids, nuclei, and particles* (2nd ed.). Wiley.
- [20] Ciaburro, G. (2017). *MATLAB for machine learning* (pp. 259–260). Packt Publishing Ltd.
- [21] Ciaburro, G. (2017). *MATLAB for machine learning* (p. 256). Packt Publishing Ltd.
- [22] O’Shea, K., & Nash, R. (2015). *An introduction to convolutional neural networks*. Recuperado de <https://doi.org/10.48550/arXiv.1511.08458>
- [23] MathWorks. (2024). *Convolutional neural network image*. Recuperado de <https://la.mathworks.com/discovery/convolutional-neural-network.html>
- [24] Wu, J. (2017). *Introduction to convolutional neural networks* (pp. 9, 11–13). Recuperado de <https://cs.nju.edu.cn/wujx/paper/CNN.pdf>
- [25] Indolia, S., Goswami, A. K., Mishra, S. P., & Asopa, P. (2018). Understanding of convolutional neural network – A deep learning approach. Recuperado de <http://www.sciencedirect.com/>
- [26] Geras, K. J., Wolfson, S., Shen, Y., Wu, N., Kim, S. G., Moy, L., & Cho, K. (2017). High-resolution breast cancer screening with multi-view deep convolutional neural networks. *arXiv preprint arXiv:1703.07047*. <https://arxiv.org/abs/1703.07047>
- [27] Pires, R., Avila, S., & Valle, E. (2019). Exploring age and quality factors in deep learning-based face recognition.
- [28] Baccouche, A., Garcia-Zapirain, B., & Elmaghraby, A. (2021). Connected-UNets: A deep learning architecture for breast mass segmentation. *npj Breast Cancer*, 7(1), 1–12. <https://doi.org/10.1038/s41523-021-00358-x>
- [29] MathWorks. (2024). *MATLAB documentation*. <https://la.mathworks.com/help/matlab/index.htmls/tid=hc/panel>
- [30] University of Essex. (s.f.). *Mammographic Image Analysis Society (MIAS) database*. Recuperado de <http://peipa.essex.ac.uk/info/mias.html>
- [31] MathWorks. (2024). *Image Labeler App*. <https://la.mathworks.com/help/vision/ref/imagelabeler-app.html>
- [32] MathWorks. (2024). *What is a Datastore*. <https://la.mathworks.com/help/matlab/import/export/what-is-a-datastore.html>
- [33] MathWorks. (2024). *MATLAB ImageDatastore*. <https://la.mathworks.com/help/matlab/ref/matlab.io.datastore.imagedatastore.html>
- [34] MathWorks. (2024). *PixelLabelDatastore*. <https://la.mathworks.com/help/vision/ref/pixellabeldatastore.html>

- [35] MathWorks. (2024). *Imresize*. <https://la.mathworks.com/help/matlab/ref/imresize.html>
- [36] MathWorks. (2024). *Pretrained Model Image*. Recuperado de https://la.mathworks.com/help/deeplearning/ug/pretrained_20b.png
- [37] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. En *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770–778). <https://doi.org/10.1109/CVPR.2016.90>
- [38] Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017). Inception-v4, Inception-ResNet and the impact of residual connections on learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 31(1). <https://doi.org/10.1609/aaai.v31i1.11231>
- [39] MathWorks. (2024). *DeepLab v3+*. <https://la.mathworks.com/help/vision/ref/deeplabv3plus.html>
- [40] MathWorks. (2024). *ImageInputLayer*. <https://la.mathworks.com/help/deeplearning/ref/nnet.cnn.layer.imageinputlayer.html>
- [41] MathWorks. (2024). *ReplaceLayer*. <https://la.mathworks.com/help/deeplearning/ref/nnet.cnn.layergraph.replacelayer.html>
- [42] MathWorks. (2024). *trainingOptions*. <https://la.mathworks.com/help/deeplearning/ref/trainingoptions.html>
- [43] MathWorks. (2024). *trainNet*. <https://la.mathworks.com/help/deeplearning/ref/trainnet.html>